

M. Nelkon & P. Parker

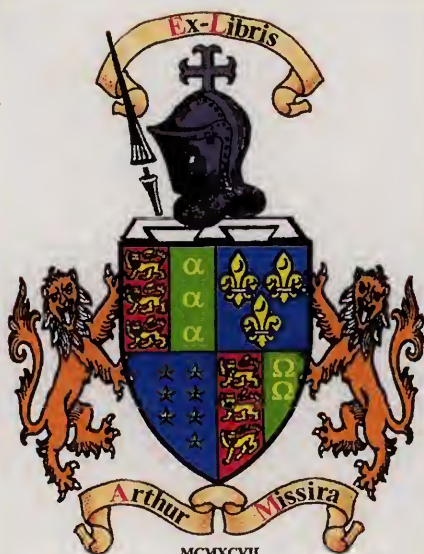
Advanced Level Physics

M. Nelkon & P. Parker

Advanced Level Physics

Third Edition With SI Units





Advanced Level Physics

Books by M. Nelkon

Published by Heinemann

ADVANCED LEVEL PRACTICAL PHYSICS (*with J. Ogborn*)

SCHOLARSHIP PHYSICS (SI)

LIGHT AND SOUND (SI)

MECHANICS AND PROPERTIES OF MATTER (SI)

PRINCIPLES OF ATOMIC PHYSICS AND ELECTRONICS (SI)

REVISION NOTES IN PHYSICS (SI)

Book I, Heat, Light, Sound

Book II. Electricity, Atomic Physics, Mechanics, Properties of Matter

GRADED EXERCISES IN PHYSICS (SI)

TEST PAPERS IN PHYSICS (SI)

REVISION BOOK IN ORDINARY LEVEL PHYSICS (SI)

ELEMENTARY PHYSICS, Book I and II (*with A. F. Abbott*)

AN INTRODUCTION TO THE MATHEMATICS OF PHYSICS
(*with J. H. Avery*)

GENERAL SCIENCE PHYSICS

ELECTRONICS AND RADIO (*with H. I. Humphreys*)

SOLUTIONS TO ADVANCED LEVEL PHYSICS QUESTIONS (SI)

SOLUTIONS TO ORDINARY LEVEL PHYSICS QUESTIONS (SI)

Published by Chatto & Windus

FUNDAMENTALS OF PHYSICS (O-Level, SI)

EXERCISES IN ORDINARY LEVEL PHYSICS (*with SI units*)

C.S.E. PHYSICS

REVISION BOOK IN C.S.E. PHYSICS

SI UNITS: AN INTRODUCTION FOR ADVANCED LEVEL

Published by Edward Arnold

ELECTRICITY AND MAGNETISM

ELECTRICITY (Advanced Level, SI)

Published by Blackie

HEAT (Advanced Level, SI)

Books by P. Parker

Published by Heinemann

INTERMEDIATE HEAT

ELECTRICITY AND MAGNETISM

Published by Arnold

ELECTRONICS

Advanced Level Physics

Third edition with SI units

M. NELKON, M.Sc.(Lond.), F.Inst.P., A.K.C.

*formerly Head of the Science Department,
William Ellis School, London*

P. PARKER, M.Sc., F.Inst.P., A.M.I.E.E.

*Late Senior Lecturer in Physics,
The City University, London*



**HEINEMANN EDUCATIONAL
BOOKS LTD · LONDON**

Heinemann Educational Books Ltd
LONDON EDINBURGH MELBOURNE TORONTO
SINGAPORE JOHANNESBURG AUCKLAND
IBADAN HONG KONG NAIROBI
NEW DELHI

ISBN 0 435 68636 4

© M. Nelkon and Mrs. P. Parker 1958, 1961, 1964, 1965
1966, 1968, 1970

First published as one volume 1958

Reprinted 1959, 1960, 1961, 1962

Second edition 1964

Reprinted with additions 1965

Reprinted with additions 1966

Reprinted with corrections 1968

Third edition (SI) 1970

Reprinted 1970

Published by
Heinemann Educational Books Ltd
48 Charles Street, London W1X 8AH
Parts 1, 2 and 4 Phototypeset by
Keypools Ltd., Golborne, Lancs.
Illustrations drawn by
Bucken Ltd., Wealdstone, Middlesex
Printed and bound in Great Britain
by Jarrold and Sons Ltd, Norwich

Preface to Third Edition

This edition covers the new syllabus of the examining boards and is written in SI units to conform to their use in all future Advanced level examinations.

The main change in the text has been in the sections in Electricity on magnetic and electric fields and their associated phenomena. In the treatment, (i) magnetic flux density or induction **B** and electric intensity **E** have been used in preference to **H** and **D**—this follows the recommendations of the 1966 report of the Association for Science Education, (ii) magnetic flux density has been defined from the relation $F = BIl$. A new section on electromagnetic waves has been added.

Other changes are as follows: *Waves*. This has now been treated generally. *Optics*. The sections on interference and diffraction have been expanded. *Sound*. An account of recording on tape and film has been added. *Heat*. The joule has been used as the unit of heat and van der Waals' equation has been discussed. *Properties of Matter*. The repulsive and attractive forces between molecules have been emphasised. *Mechanics*. There are further discussions on angular momentum and the dynamics of a rigid body. Throughout the text, worked examples in SI units have been given to assist the student and the exercises at the end of chapters contain recent questions from examining boards. It is hoped that this SI edition will continue to assist students of Advanced level standard.

We are very much indebted to M. V. Detheridge, Woodhouse Grammar School, London, for his valuable co-operation in the writing and the preparation of the new electricity sections and for his generous assistance with this edition. We are also grateful to Rev. M. D. Phillips, O.S.B., Ampleforth College, York; S. S. Alexander, Woodhouse Grammar School, London; C. A. Boyle, William Ellis School, London; S. Freake, Queen's College, Cambridge, and R. P. T. Hills, St. John's College, Cambridge, for reading parts of the work; and to Prof. M. L. McGlashan, Exeter University, and M. Sayer, Chetham's Hospital School, Manchester, for advice on SI units.

I am grateful to the following for permission to include photographs in this book. To the Head of the Physics Department, the City University, London, for Newton rings, Fresnel biprism interference bands, Diffraction rings and Diffraction bands; to the late Sir J. J. Thomson for Positive Rays photographs; to the National Chemical Laboratory, for X-Ray diffraction rings; to Lord Blackett of the Imperial College of Science and Technology, for Transmutation of Nitrogen; to Professor George Thomson and the Science Museum, for

Electron diffraction rings; and finally to the United Kingdom Atomic Energy Authority for Van der Graff Electrostatic Generator and Nuclear Research Reactor.

Thanks are due to the following Examining Boards for their kind permission to translate numerical quantities in past questions to SI units; the translation is the sole responsibility of the author:

London University School Examinations (*L.*),
Oxford and Cambridge Schools Examination Board (*O. & C.*),
Joint Matriculation Board (*N.*),
Cambridge Local Examinations Syndicate (*C.*),
Oxford Delegacy of Local Examinations (*O.*).

1970

M.N.

Preface to First Edition

This text-book is designed for Advanced Level students of Physics, and covers Mechanics and Properties of Matter, Heat, Optics, and Sound. Electricity and Atomic Physics to that standard. It is based on the experience gained over many years of teaching and lecturing to a wide variety of students in schools and polytechnics.

In the treatment, an Ordinary Level knowledge of the subject is assumed. We have aimed at presenting the physical aspect of topics as much as possible, and then at providing the mathematical arguments and formulae necessary for a thorough understanding. Historical details have also been given to provide a balanced perspective of the subject. As a help to the student, numerous worked examples from past examination papers have been included in the text.

It is possible here to mention only a few points borne in mind by the authors. In Mechanics and Properties of Matter, the theory of dimensions has been utilized where the mathematics is difficult, as in the subject of viscosity, and the "excess pressure" formula has been extensively used in the treatment of surface tension. In Heat, the kinetic theory of gases has been fully discussed, and the experiments of Joule and Andrews have been presented in detail. The constant value of $n \sin i$ has been emphasized in refraction at plane surfaces. In Optics, there is a full treatment of optical treatment of optical instruments, and accounts of interference, diffraction and polarization. In Sound, the physical principles of stationary waves, and their applications to pipes and strings, have been given prominence. Finally, in Electricity the electron and ion have been used extensively to produce explanations of phenomena in electrostatics, electromagnetism, electrolysis and atomic physics; the concept of e.m.f. has been linked at the outset with energy; and there are accounts of measurements and instruments.

We acknowledge our gratitude to the following for their kindness in reading sections of the work before the complete volume was compiled: Mr. J. H. Avery, Stockport Grammar School; Dr. J. Duffey, formerly of Watford Technical College; Mr. J. Newton, The City University, London; Mr. A. W. K. Ingram, Lawrence Sheriff School, Rugby; Mr. O. C. Gay, College of Technology, Hull; Mr. T. N. Littledale, Gunnersbury Grammar School; Mr. C. R. Ensor, Downside School, Bath; Mr. L. S. Powell, Garnett College, London; Dr. D. W. Stops, The City University, London; and Professor H. T. Flint, formerly of London University.

Preface to Second Edition

In this edition I have added an introduction to Atomic Structure, which covers the Advanced level syllabus on this topic. I am particularly indebted to Mr. J. Yarwood, M.Sc., F.Inst.P., head of the physics and mathematics department, Regent Street Polytechnic, London, for reading this section and for valuable advice, and to Prof. L. Pincherle, Bedford College, London University, for his kind assistance in parts of the text.

I am also indebted to G. Ullyott, Charterhouse School and L. G. Mead, Wellington School, Somerset, for their helpful comments on dynamics and optics respectively.

Contents

PART ONE: MECHANICS AND PROPERTIES OF MATTER

	<i>Page</i>
1 Dynamics	1
2 Circular motion, S.H.M. Gravitation	36
3 Rotation of rigid bodies	75
4 Static bodies, Fluids	96
5 Surface tension	125
6 Elasticity	153
7 Solid Friction. Viscosity	171

PART TWO: HEAT

8 Introduction	189
9 Calorimetry	199
10 Gases	220
11 Thermal expansion	269
12 Changes of State	294
13 Transfer of heat	330
14 Thermometry and pyrometry	366

PART THREE: OPTICS AND SOUND

15 Optics: Introduction	387
16 Reflection at plane surfaces	391
17 Reflection at curved surfaces	402
18 Refraction at plane surfaces	420
19 Refraction through prisms	441
20 Dispersion. Spectra	454
21 Refraction through lenses	471
22 Defects of vision. Defects of lenses	508
23 Optical instruments	525
24 Velocity of light. Photometry	551
25 Oscillations and waves. Sound waves	577
26 Characteristics, properties, velocity of sound	606
27 Vibrations in pipes, strings, rods	640
28 Optics: Wave theory of light	675
29 Interference, diffraction, polarization	687

PART FOUR: ELECTRICITY AND ATOMIC PHYSICS

	<i>Page</i>
30 Electrostatics	731
31 Capacitors	765
32 Current electricity. Resistance. Power.	785
33 Ohm's law applications. Measurement. Networks.	807
34 Chemical effect of current	844
35 Magnetic field. Force on conductor	874
36 Electromagnetic induction	895
37 Magnetic fields due to conductors	931
38 Magnetic properties of material	948
39 A.C. circuits. Electromagnetic waves	967
40 Electrons. Motion in fields	993
41 Valves. C.R.O. Junction diode. Transistors	1009
42 Radioactivity. The nucleus	1040
43 X-rays. Photoelectricity. Energy levels	1068
Summary of C.G.S. and SI Units	1095
Answers to exercises	1101
Index	1107

PART ONE

Mechanics and Properties of Matter

chapter one

Dynamics

Motion in a Straight Line. Velocity

If a car travels steadily in a constant direction and covers a distance s in a time t , then its *velocity* in that direction = s/t . If the car does not travel steadily, then s/t is its average velocity, and

$$\text{distance } s = \text{average velocity} \times t.$$

We are here concerned with motion in a constant direction. The term 'displacement' is given to the distance moved in a constant direction, for example, from L to C in Fig. 1.1 (i). Velocity may therefore be defined as the *rate of change of displacement*.

Velocity can be expressed in *centimetres per second* (cm/s or cm s^{-1}) or *metres per second* (m/s or m s^{-1}) or *kilometres per hour* (km/h or km h^{-1}). By calculations, $36 \text{ km h}^{-1} = 10 \text{ m s}^{-1}$. It should be noted that complete information is provided for a velocity by stating its direction in addition to its magnitude, as explained shortly.

If an object moving in a straight line travels equal distances in equal times, no matter how small these distances may be, the object is said to be moving with *uniform velocity*. The velocity of a falling stone increases continuously, and so is a *non-uniform velocity*.

If, at any point of a journey, Δs is the small change in displacement in a small time Δt , the velocity v is given by $v = \Delta s / \Delta t$. In the limit, using calculus notation,

$$v = \frac{ds}{dt}$$

Vectors

Displacement and *velocity* are examples of a class of quantities called *vectors* which have both magnitude and direction. They may therefore be represented to scale by a line drawn in a particular direction. Thus

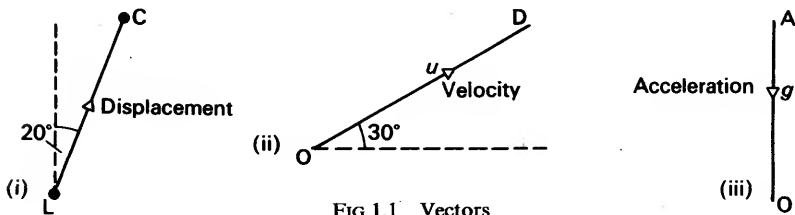


FIG 1.1 Vectors

Cambridge is 80 km from London in a direction 20° E. of N. We can therefore represent the displacement between the cities in magnitude

and direction by a straight line LC 4 cm long 20° E. of N., where 1 cm represents 20 km, Fig. 1.1 (i). Similarly, we can represent the velocity u of a ball initially thrown at an angle of 30° to the horizontal by a straight line OD drawn to scale in the direction of the velocity u , the arrow on the line showing the direction, Fig. 1.1 (ii). The acceleration due to gravity, g , is always represented by a straight line AO to scale drawn vertically downwards, since this is the direction of the acceleration, Fig. 1.1 (iii). We shall see later that 'force' and 'momentum' are other examples of vectors.

Speed and Velocity

A car moving along a winding road or a circular track at 80 km h^{-1} is said to have a *speed* of 80 km h^{-1} . 'Speed' is a quantity which has no direction but only magnitude, like 'mass' or 'density' or 'temperature'. These quantities are called *scalars*.

The distinction between speed and velocity can be made clear by reference to a car moving round a circular track at 80 km h^{-1} say. Fig. 1.2. At every point on the track the *speed* is the same—it is 80 km h^{-1} .

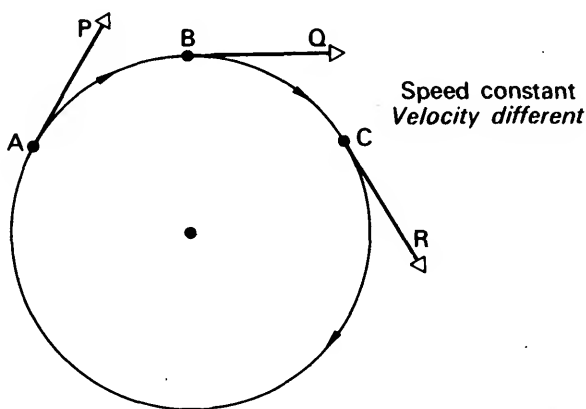


FIG. 1.2. Velocity and speed

At every point, however, the *velocity* is different. At A, B or C, for example, the velocity is in the direction of the particular tangent, AP, BQ or CR, so that even though the magnitudes are the same, the three velocities are all different because they point in different directions. Generally, vector quantities can be represented by a line drawn in the direction of the vector and whose length represents its magnitude.

Distance-Time Curve

When the displacement, or distance, s of a moving car from some fixed point is plotted against the time t , a *distance-time* ($s-t$) curve of

the motion is obtained. The velocity of the car at any instant is given by the change in distance per second at that instant. At E, for example, if the change in distance s is Δs and this change is made in a time Δt ,

$$\text{velocity at E} = \frac{\Delta s}{\Delta t}.$$

In the limit, then, when Δt approaches zero, the velocity at E becomes equal to the *gradient of the tangent to the curve at E*. Using calculus notation, $\Delta s/\Delta t$ then becomes equal to ds/dt (p. 1).

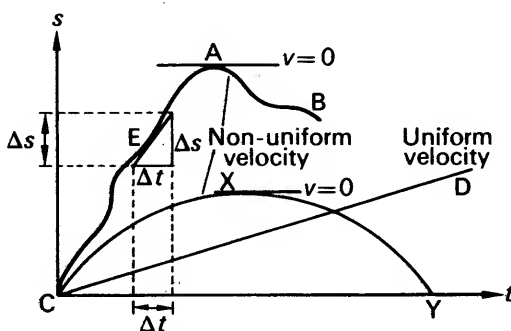


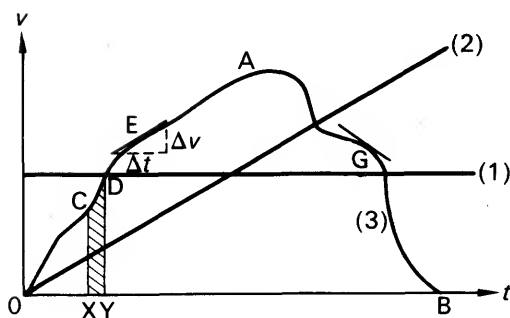
FIG. 1.3 Displacement (s)-time (t) curves

If the distance-time curve is a straight line CD, the gradient is constant at all points; it therefore follows that the car is moving with a *uniform* velocity, Fig. 1.3. If the distance-time curve is a curve CAB, the gradient varies at different points. The car then moves with non-uniform velocity. We may deduce that the velocity is zero at the instant corresponding to A, since the gradient at A to the curve CAB is zero.

When a ball is thrown upwards, the height s reached at any instant t is given by $s = ut - \frac{1}{2}gt^2$, where u is the initial velocity and g is the constant equal to the acceleration due to gravity (p. 8). The graph of s against t is represented by the parabolic curve CXY in Fig. 1.3; the gradient at X is zero, illustrating that the velocity of the ball at its maximum height is zero.

Velocity-Time Curves

When the velocity of a moving train is plotted against the time, a 'velocity-time (v - t) curve' is obtained. Useful information can be deduced from this curve, as we shall see shortly. If the velocity is uniform, the velocity-time graph is a straight line parallel to the time-axis, as shown by line (1) in Fig. 1.4. If the train accelerates uniformly from rest, the velocity-time graph is a straight line, line (2), inclined to the time-axis. If the acceleration is not uniform, the velocity-time graph is curved.

FIG. 1.4 Velocity (v)-time (t) curves

In Fig. 1.4, the velocity-time graph OAB represents the velocity of a train starting from rest which reaches a maximum velocity at A, and then comes to rest at the time corresponding to B; the acceleration and retardation are both not uniform in this case.

Acceleration is the 'rate of change of velocity', i.e. the change of velocity per second. *The acceleration of the train at any instant is given by the gradient to the velocity-time graph at that instant, as at E.* At the peak point A of the curve OAB the gradient is zero, i.e., the acceleration is then zero. At any point, such as G, between A, B the gradient to the curve is negative, i.e., the train undergoes retardation.

The gradient to the curve at any point such as E is given by :

$$\frac{\text{velocity change}}{\text{time}} = \frac{\Delta v}{\Delta t}$$

where Δv represents a small change in v in a small time Δt . In the limit, the ratio $\Delta v/\Delta t$ becomes dv/dt , using calculus notation.

Area Between Velocity-Time Graph and Time-Axis

Consider again the velocity-time graph OAB, and suppose the velocity increases in a very small time-interval XY from a value represented by XC to a value represented by YD, Fig. 1.4. Since the small distance travelled = average velocity \times time XY, the distance travelled is represented by the *area* between the curve CD and the time-axis, shown shaded in Fig. 1.4. By considering every small time-interval between OB in the same way, it follows that *the total distance travelled by the train in the time OB is given by the area between the velocity-time graph and the time-axis.* This result applies to any velocity-time graph, whatever its shape.

Fig. 1.5 illustrates the velocity-time graph AB of an object moving with uniform acceleration a from an initial velocity u . From above, the distance s travelled in a time t or OC is equivalent to the area OABC. The area OADC = $u \cdot t$. The area of the triangle ABD =

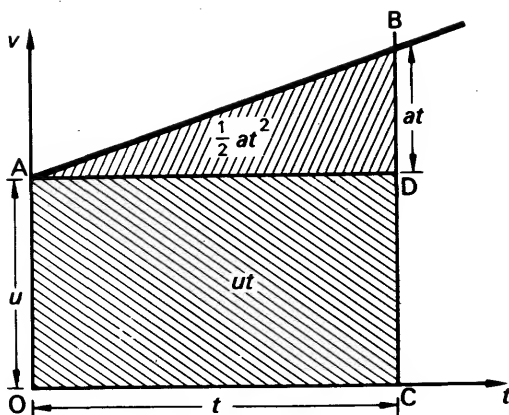


FIG. 1.5 Uniform acceleration

$\frac{1}{2}AD \cdot BD = \frac{1}{2}t \cdot BD$. Now $BD =$ the increase in velocity in a time $t = at$. Hence area of triangle $ABD = \frac{1}{2}t \cdot at = \frac{1}{2}at^2$

$$\therefore \text{total area } OABC = s = ut + \frac{1}{2}at^2.$$

This result is also deduced on p. 6.

Acceleration

The *acceleration* of a moving object at an instant is the *rate of change of its velocity* at that instant. In the case of a train accelerating steadily from 36 km h^{-1} (10 m s^{-1}) to 54 km h^{-1} (15 m s^{-1}) in 10 second, the uniform acceleration

$$= (54 - 36) \text{ km h}^{-1} \div 10 \text{ seconds} = 1.8 \text{ km h}^{-1} \text{ per second,}$$

or

$$(15 - 10) \text{ m s}^{-1} \div 10 \text{ seconds} = 0.5 \text{ m s}^{-1} \text{ per second.}$$

Since the time element (second) is repeated twice in the latter case, the acceleration is usually given as 0.5 m s^{-2} . Another unit of acceleration is ' cm s^{-2} '. In terms of the calculus, the acceleration a of a moving object is given by

$$a = \frac{dv}{dt}$$

where dv/dt is the velocity change per second.

Distance Travelled with Uniform Acceleration. Equations of Motion

If the velocity changes by equal amounts in equal times, no matter how small the time-intervals may be, the acceleration is said to be *uniform*. Suppose that the velocity of an object moving in a straight

line with uniform acceleration a increases from a value u to a value v in a time t . Then, from the definition of acceleration,

$$a = \frac{v-u}{t},$$

from which

$$v = u + at \quad (1)$$

Suppose an object with a velocity u accelerates with a uniform acceleration a for a time t and attains a velocity v . The distance s travelled by the object in the time t is given by

$$s = \text{average velocity} \times t$$

$$= \frac{1}{2}(u+v) \times t$$

But

$$v = u + at$$

$$\therefore s = \frac{1}{2}(u + u + at)t$$

$$\therefore s = ut + \frac{1}{2}at^2 \quad (2)$$

If we eliminate t by substituting $t = (v-u)/a$ from (1) in (2), we obtain, on simplifying,

$$v^2 = u^2 + 2as \quad (3)$$

Equations (1), (2), (3) are the equations of motion of an object moving in a straight line with uniform acceleration. When an object undergoes a uniform *retardation*, for example when brakes are applied to a car, a has a *negative* value.

EXAMPLES

1. A car moving with a velocity of 54 km h^{-1} accelerates uniformly at the rate of 2 m s^{-2} . Calculate the distance travelled from the place where acceleration began to that where the velocity reaches 72 km h^{-1} , and the time taken to cover this distance.

(i) $54 \text{ km h}^{-1} = 15 \text{ m s}^{-1}$, $72 \text{ km h}^{-1} = 20 \text{ m s}^{-1}$, acceleration $a = 2 \text{ m s}^{-2}$.

Using

$$v^2 = u^2 + 2as,$$

$$\therefore 20^2 = 15^2 + 2 \times 2 \times s$$

$$\therefore s = \frac{20^2 - 15^2}{2 \times 2} = 43\frac{3}{4} \text{ m.}$$

(ii) Using

$$v = u + at$$

$$\therefore 20 = 15 + 2t$$

$$\therefore t = \frac{20 - 15}{2} = 2.5 \text{ s.}$$

2. A train travelling at 72 km h^{-1} undergoes a uniform retardation of 2 m s^{-2} when brakes are applied. Find the time taken to come to rest and the distance travelled from the place where the brakes were applied.

(i) $72 \text{ km h}^{-1} = 20 \text{ m s}^{-1}$, and $a = -2 \text{ m s}^{-2}$, $v = 0$.

Using

$$v = u + at$$

$$\therefore 0 = 20 - 2t$$

$$\therefore t = 10 \text{ s}$$

(ii) The distance, s , $= ut + \frac{1}{2}at^2$.

$$= 20 \times 10 - \frac{1}{2} \times 2 \times 10^2 = 100 \text{ m.}$$

Motion Under Gravity

When an object falls to the ground under the action of gravity, experiment shows that the object has a constant or uniform acceleration of about 980 cm s^{-2} , while it is falling (see p. 49). In SI units this is 9.8 m s^{-2} or 10 m s^{-2} approximately. The numerical value of this acceleration is usually denoted by the symbol g . Suppose that an object is dropped from a height of 20 m above the ground. Then the initial velocity $u = 0$, and the acceleration $a = g = 10 \text{ m s}^{-2}$ (approx). Substituting in $s = ut + \frac{1}{2}at^2$, the distance fallen s in metres is calculated from

$$s = \frac{1}{2}gt^2 = 5t^2.$$

When the object reaches the ground, $s = 20 \text{ m}$.

$$\therefore 20 = 5t^2, \text{ or } t = 2 \text{ s}$$

Thus the object takes 2 seconds to reach the ground.

If a cricket-ball is thrown vertically upwards, it slows down owing to the attraction of the earth. The ball is thus retarded. The magnitude of the retardation is 9.8 m s^{-2} , or g . Mathematically, a retardation can be regarded as a negative acceleration in the direction along which the object is moving; and hence $a = -9.8 \text{ m s}^{-2}$ in this case.

Suppose the ball was thrown straight up with an initial velocity, u , of 30 m s^{-1} . The time taken to reach the top of its motion can be obtained from the equation $v = u + at$. The velocity, v , at the top is zero; and since $u = 30 \text{ m s}^{-1}$ and $a = -9.8$ or 10 m s^{-2} (approx), we have

$$0 = 30 - 10t.$$

$$\therefore t = \frac{30}{10} = 3 \text{ s.}$$

The highest distance reached is thus given by

$$\begin{aligned} s &= ut + \frac{1}{2}at^2 \\ &= 30 \times 3 - 5 \times 3^2 = 45 \text{ m.} \end{aligned}$$

Resultant. Components

If a boy is running along the deck of a ship in a direction OA, and the

ship is moving in a different direction OB, the boy will move relatively to the sea along a direction OC, between OA and OB, Fig. 1.6 (i). Now in one second the boat moves from O to B, where OB represents the velocity of the boat, a vector quantity, in magnitude and direction. The boy moves from O to A in the same time, where OA represents the velocity of the boy in magnitude and direction. Thus in one second the net effect relative to the sea is that the boy moves from O to C. It can now be seen that if lines OA, OB are drawn to represent in magnitude and direction the respective velocities of the boy and the ship, the magnitude and direction of the *resultant* velocity of the boy is represented by the diagonal OC of the completed parallelogram having OA, OB as two of its sides; OACB is known as a *parallelogram of velocities*. Conversely, a velocity represented completely by OC can be regarded as having an 'effective part', or *component* represented by OA, and another component represented by OB.

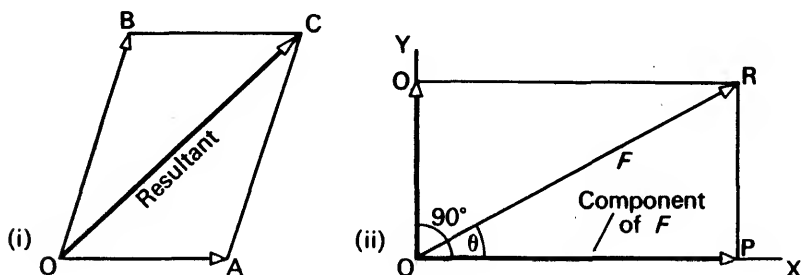


FIG. 1.6. Resultant and component.

In practice, we often require to find the component of a vector quantity in a certain direction. Suppose OR represents the vector F , and OX is the direction, Fig. 1.6 (ii). If we complete the parallelogram OQRP by drawing a perpendicular RP from R to OX, and a perpendicular RQ from R to OY, where OY is perpendicular to OX, we can see that OP, OQ represent the components of F along OX, OY respectively. Now the component OQ has no effect in a perpendicular direction; consequently OP represents the total effect of F along the direction OX. OP is called the 'resolved component' in this direction. If θ is the angle ROX, then, since triangle OPR has a right angle at P,

$$OP = OR \cos \theta = F \cos \theta \quad (4)$$

Components of g

The acceleration due to gravity, g , acts vertically downwards. In free fall, an object has an acceleration g . An object sliding freely down an inclined plane, however, has an acceleration due to gravity equal to the component of g down the plane. If it is inclined at 60° to the vertical, the acceleration down the plane is then $g \cos 60^\circ$ or $9.8 \cos 60^\circ \text{ m s}^{-2}$, which is 4.9 m s^{-2} .

Consider an object O thrown forward from the top of a cliff OA

with a horizontal velocity u of 15 m s^{-1} . Fig. 1.7. Since u is horizontal, it has no component in a *vertical* direction. Similarly, since g acts vertically, it has no component in a *horizontal* direction.

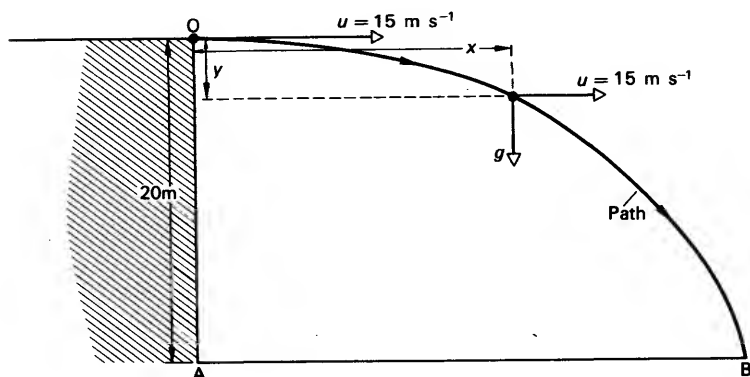


FIG. 1.7 Motion under gravity

We may thus treat vertical and horizontal motion independently. Consider the vertical motion from O. If OA is 20 m, the ball has an initial vertical velocity of zero and a vertical acceleration of g , which is 9.8 m s^{-2} (10 m s^{-2} approximately). Thus, from $s = ut + \frac{1}{2}at^2$, the time t to reach the bottom of the cliff is given, using $g = 10 \text{ m s}^{-2}$, by

$$20 = \frac{1}{2} \cdot 10 \cdot t^2 = 5t^2, \text{ or } t = 2 \text{ s.}$$

So far as the horizontal motion is concerned, the ball continues to move forward with a constant velocity of 15 m s^{-1} since g has no component horizontally. In 2 seconds, therefore,

horizontal distance AB = distance from cliff = $15 \times 2 = 30 \text{ m}$.

Generally, in a time t the ball falls a vertical distance, y say, from O given by $y = \frac{1}{2}gt^2$. In the same time the ball travels a horizontal distance, x say, from O given by $x = ut$, where u is the velocity of 15 m s^{-1} . If t is eliminated by using $t = x/u$ in $y = \frac{1}{2}gt^2$, we obtain $y = gx^2/2u$. This is the equation of a *parabola*. It is the path OB in Fig. 1.7.

Addition of Vectors

Suppose a ship is travelling due east at 30 km h^{-1} and a boy runs across the deck in a north-west direction at 6 km h^{-1} , Fig. 1.8 (i). We

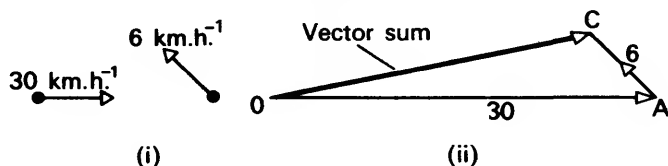


FIG. 1.8 Addition of vectors

can find the velocity and direction of the boy relative to the sea by adding the two velocities. Since velocity is a vector quantity, we draw a line OA to represent 30 km h^{-1} in magnitude and direction, and then, from the end of A , draw a line AC to represent 6 km h^{-1} in magnitude and direction, Fig. 1.8 (ii). The sum, or resultant, of the velocities is now represented by the line OC in magnitude and direction, because a distance moved in one second by the ship (represented by OA) together with a distance moved in one second by the boy (represented by AC) is equivalent to a movement of the boy from O to C relative to the sea.

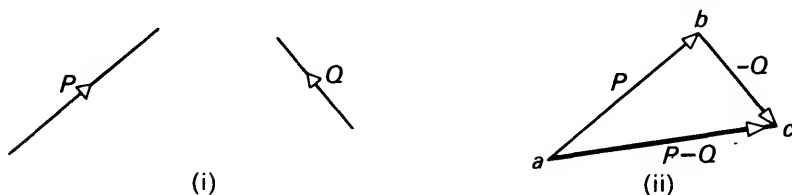


FIG. 1.9 Subtraction of velocities

In other words, the difference between the vectors \vec{P} , \vec{Q} in Fig. 1.9 (i) is the *sum* of the vectors \vec{P} and $(-\vec{Q})$. Now $(-\vec{Q})$ is a vector drawn exactly equal and opposite to the vector \vec{Q} . We therefore draw ab to represent \vec{P} completely, and then draw bc to represent $(-\vec{Q})$ completely, Fig. 1.9 (ii). Then $\vec{P} + (-\vec{Q}) =$ the vector represented by $ac = \vec{P} - \vec{Q}$.

Relative Velocity and Relative Acceleration

If a car A travelling at 50 km h^{-1} is moving in the same direction as another car B travelling at 60 km h^{-1} , the *relative velocity* of B to A = $60 - 50 = 10 \text{ km h}^{-1}$. If, however, the cars are travelling in opposite directions, the relative velocity of B to A = $60 - (-50) = 110 \text{ km h}^{-1}$.

Suppose that a car X is travelling with a velocity v along a road 30° east of north, and a car Y is travelling with a velocity u along a road due east, Fig. 1.10 (i). Since 'velocity' has direction as well as magnitude, i.e., 'velocity' is a vector quantity (p. 1), we cannot subtract u and v numerically to find the relative velocity. We must adopt a method which takes into account the direction as well as the magnitude of the velocities, i.e., a vector subtraction is required.

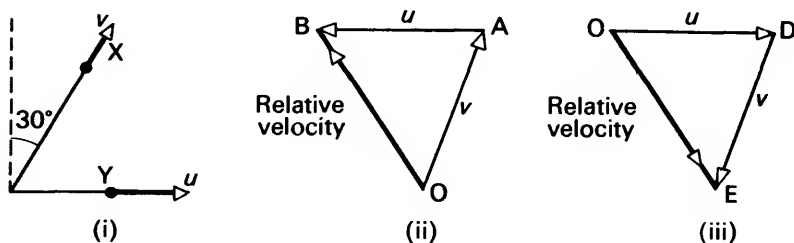


FIG. 1.10. Relative velocity.

The velocity of X relative to Y $= \vec{v} - u = \vec{v} + (-\vec{u})$. Suppose OA represents the velocity, v , of X in magnitude and direction, Fig. 1.10 (ii). Since Y is travelling due east, a velocity AB numerically equal to u but in the due west direction represents the vector $(-\vec{u})$. The vector sum of OA and AB is OB from p. 0, which therefore represents in magnitude and direction the velocity of X relative to Y. By drawing an accurate diagram of the two velocities, OB can be found.

The velocity of Y relative to X $= \vec{u} - \vec{v} = \vec{u} + (-\vec{v})$, and can be found by a similar method. In this case, OD represents the velocity, u , of Y in magnitude and direction, while DE represents the vector $(-\vec{v})$, which it is drawn numerically equal to v but in the opposite direction, Fig. 1.10 (iii). The vector sum of OD and DE is OE, which therefore represents the velocity of Y relative to X in magnitude and direction.

When two objects P, Q are each accelerating, the acceleration of P relative to Q = acceleration of P - acceleration of Q. Since 'acceleration' is a vector quantity, the relative acceleration must be found by vector subtraction, as for the case of relative velocity.

EXAMPLE

Explain the difference between a scalar and a vector quantity.

What is meant by the relative velocity of one body with respect to another? Two ships are 10 km apart on a line running S. to N. The one farther north is steaming W. at 20 km h^{-1} . The other is steaming N. at 20 km h^{-1} . What is their distance of closest approach and how long do they take to reach it? (C.)

Suppose the two ships are at X, Y, moving with velocities u, v respectively, each 20 km h^{-1} Fig. 1.11 (i). The velocity of Y relative to X $= \vec{v} - \vec{u} = \vec{v} + (-\vec{u})$. We therefore draw OA to represent \vec{v} (20) and add to it AB, which represents $(-\vec{u})$, Fig. 1.11 (ii). The relative velocity is then represented by OB.

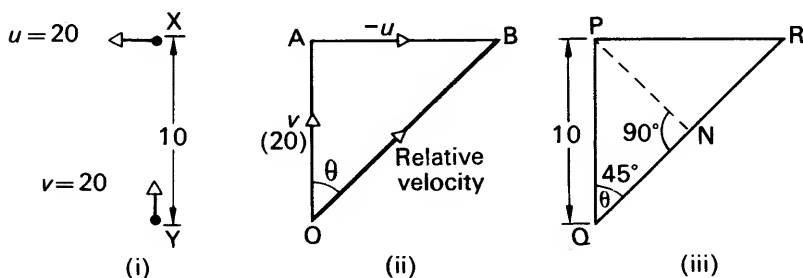


FIG. 1.11 Example

Since OAB is a right-angled triangle,

$$OB = \sqrt{OA^2 + AB^2} = \sqrt{20^2 + 20^2} = 28.28 = 28.3 \text{ km h}^{-1} \quad (i)$$

Also, $\tan \theta = \frac{AB}{OA} = \frac{20}{20} = 1$, i.e., $\theta = 45^\circ$ (ii)

Thus the ship Y will move along a direction QR relative to the ship X, where QR is at 45° to PQ, the north-south direction, Fig. 1.11(iii). If $PQ = 10$ km, the distance of closest approach is PN, where PN is the perpendicular from P to QR.

$$\therefore PN = PQ \sin 45^\circ = 10 \sin 45^\circ = 7.07 \text{ km.}$$

The distance $QN = 10 \cos 45^\circ = 7.07$ km. Since, from (i), the relative velocity is 28.28 km h^{-1} , it follows that

$$\text{time to reach N} = \frac{7.07}{28.28} = \frac{1}{4} \text{ hour.}$$

LAWS OF MOTION. FORCE AND MOMENTUM

Newton's Laws of Motion

In 1686 SIR ISAAC NEWTON published a work called *Principia*, in which he expounded the Laws of Mechanics. He formulated in the book three 'laws of motion':

Law I. *Every body continues in its state of rest or uniform motion in a straight line, unless impressed forces act on it.*

Law II. *The change of momentum per unit time is proportional to the impressed force, and takes place in the direction of the straight line along which the force acts.*

Law III. *Action and reaction are always equal and opposite.*

These laws cannot be proved in a formal way; we believe they are correct because all the theoretical results obtained by assuming their truth agree with the experimental observations, as for example in astronomy (p. 58).

Inertia. Mass

Newton's first law expresses the idea of **inertia**. The inertia of a body is its reluctance to start moving, and its reluctance to stop once it has begun moving. Thus an object at rest begins to move only when it is pushed or pulled, i.e., when a *force* acts on it. An object O moving in a

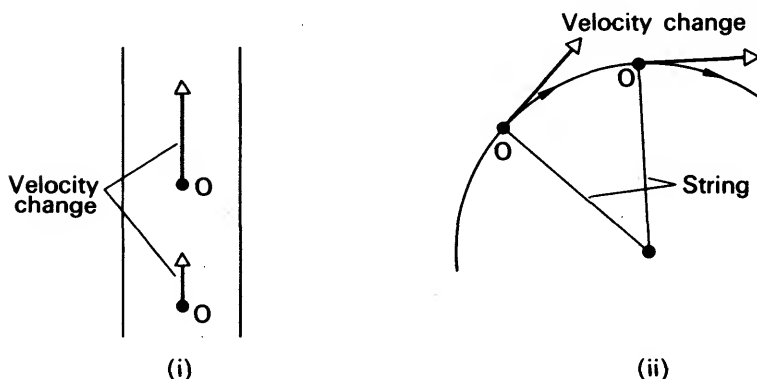


FIG. 1.12 Velocity changes

straight line with constant velocity will change its direction or move faster only if a new force acts on it. Fig. 1.12 (i). This can be demonstrated by a puck moving on a cushion of gas on a smooth level sheet of glass. As the puck slides over the glass, photographs taken at successive equal times by a stroboscopic method show that the motion is practically that of uniform velocity. Passengers in a bus or car are jerked forward when the vehicle stops suddenly. They continue in their state of motion until brought to rest by friction or collision. The use of safety belts reduces the shock.

Fig. 1.12 (ii) illustrates a velocity change when an object O is whirled at constant speed by a string. This time the magnitude of the velocity v is constant but its direction changes.

'Mass' is a measure of the inertia of a body. If an object changes its direction or its velocity slightly when a large force acts on it, its inertial mass is high. The mass of an object is constant all over the world; it is the same on the earth as on the moon. Mass is measured in kilogrammes (kg) or grammes (g) by means of a chemical balance, where it is compared with standard masses based on the International Prototype Kilogramme (see also p. 14).

Force. The newton

When an object X is moving it is said to have an amount of *momentum* given, by definition, by

$$\text{momentum} = \text{mass of } X \times \text{velocity} \quad (1)$$

Thus an object of mass 20 kg moving with a velocity of 10 m s^{-1} has a momentum of 200 kg m s^{-1} . If another object collides with X its velocity alters, and thus the momentum of X alters. From Newton's second law, a *force* acts on X which is equal to the change in momentum per second.

Thus if F is the magnitude of a force acting on a constant mass m ,

$$F \propto m \times \text{change of velocity per second}$$

$$\therefore F \propto ma,$$

where a is the *acceleration* produced by the force, by definition of a .

$$\therefore F = kma \quad (2)$$

where k is a constant.

With SI units, the **newton** (N) is the unit of force. It is defined as the force which gives a mass of 1 kilogramme an acceleration of 1 metre s^{-2} . Substituting $F = 1 \text{ N}$, $m = 1 \text{ kg}$ and $a = 1 \text{ m s}^{-2}$ in the expression for F in (i), we obtain $k = 1$. Hence, with units as stated, $k = 1$.

$$\therefore F = ma,$$

which is a standard equation in dynamics. Thus if a mass of 200 g is acted upon by a force F which produces an acceleration a of 4 m s^{-2} , then, since $m = 200 \text{ g} = 0.2 \text{ kg}$,

$$F = ma = 0.2(\text{kg}) \times 4(\text{m s}^{-2}) = 0.8 \text{ N}.$$

C.g.s. units of force

The *dyne* is the unit of force in the centimetre-gramme-second system; it is defined as the force acting on a mass of 1 gramme which gives it an acceleration of 1 cm s^{-2} . The equation $F = ma$ also applies when m is in grammes, a is in cm s^{-2} , and F is in dynes. Thus if a force of 10000 dynes acts on a mass of 200 g, the acceleration a is given by

$$F = 10000 = 200 \times a, \quad \text{or} \quad a = 50 \text{ cm s}^{-2}.$$

Suppose $m = 1 \text{ kg} = 1000 \text{ g}$, $a = 1 \text{ m s}^{-2} = 100 \text{ cm s}^{-2}$. Then, the force F is given by

$$F = ma = 1000 \times 100 \text{ dynes} = 10^5 \text{ dynes}.$$

But the force acting on a mass of 1 kg which gives it an acceleration of 1 m s^{-2} is the *newton*, N. Hence

$$1 \text{ N} = 10^5 \text{ dynes}$$

Weight. Relation between newton, kgf and dyne, gf

The *weight* of an object is defined as the *force* acting on it due to gravity; the weight of an object can hence be measured by attaching it to a spring-balance and noting the extension, as the latter is proportional to the force acting on it (p. 50).

Suppose the weight of an object of mass m is denoted by W . If the object is released so that it falls to the ground, its acceleration is g . Now $F = ma$. Consequently the force acting on it, i.e., its weight, is given by

$$W = mg.$$

If the mass is 1 kg, then, since $g = 9.8 \text{ m s}^{-2}$, the weight $W = 1 \times 9.8 = 9.8 \text{ N}$ (newton). The force due to gravity on a mass of 1 kg where g has the value 9.80665 m s^{-2} is called a 1 *kilogramme force* or 1 kgf (this is roughly equal to 1 kilogramme weight or 1 kg wt, which depends on the value of g and thus varies from place to place). Hence it follows that

$$1 \text{ kgf} = 9.8 \text{ N} = 10 \text{ N approximately}.$$

A weight of 5 kgf is thus about 50 N. Further, $1 \text{ N} = \frac{1}{10} \text{ kgf approx} = 100 \text{ gf}$. The weight of an apple is about 1 newton.

The weight of a mass of 1 gramme is called *gramme-force* (1 gf); it was formerly called '1 gramme wt'. From $F = ma$, it follows that

$$1 \text{ gf} = 1 \times 980 = 980 \text{ dynes}.$$

since $g = 980 \text{ cm s}^{-2}$ (approx).

The reader should note carefully the difference between the 'kilogramme' and the 'kilogramme force'; the former is a *mass* and is therefore constant all over the universe, whereas the kilogramme force is a *force* whose magnitude depends on the value of g . The acceleration due to gravity, g , depends on the distance of the place considered from the centre of the earth; it is slightly greater at the poles than at the

equator, since the earth is not perfectly spherical (see p. 41). It therefore follows that the weight of an object differs in different parts of the world. On the moon, which is smaller than the earth and has a smaller density, an object would weigh about one-sixth of its weight on the earth.

The relation $F = ma$ can be verified by using a ticker-tape and timer to measure the acceleration of a moving trolley. Details are given in a more basic text, such as *Fundamentals of Physics* (Chatto and Windus) by the author.

The following examples illustrate the application of $F = ma$. It should be carefully noted that (i) F represents the *resultant* force on the object of mass m , (ii) F must be expressed in the appropriate units of a 'force' and m in the corresponding units of a 'mass'.

EXAMPLES

1. A force of 20 kgf pulls a sledge of mass 50 kg and overcomes a constant frictional force of 4 kgf. What is the acceleration of the sledge?

$$\text{Resultant force, } F, = 20 \text{ kgf} - 4 \text{ kgf} = 16 \text{ kgf.}$$

To change this to units of newtons, use $1 \text{ kgf} = 9.8 \text{ N} = 10 \text{ N approx.}$

$$\therefore 16 \text{ kgf} = 160 \text{ N approx.}$$

From $F = ma$,

$$\therefore 160 = 50 \times a$$

$$\therefore a = 3.2 \text{ m s}^{-2}.$$

2. An object of mass 2.00 kg is attached to the hook of a spring-balance, and the latter is suspended vertically from the roof of a lift. What is the reading on the spring-balance when the lift is (i) ascending with an acceleration of 20 cm s^{-2} , (ii) descending with an acceleration of 10 cm s^{-2} , (iii) ascending with a uniform velocity of 15 cm s^{-1} .

Suppose T is the tension (force) in the spring-balance in kgf.

(i) The object is acted upon two forces: (a) The tension T kgf in the spring-balance, which acts upwards, (b) its weight, 2 kgf, which acts downwards. Since the object moves upwards, T is greater than 2 kgf. Hence the net force, F , acting on the object $= (T - 2) \text{ kgf} = (T - 2) \times 10 \text{ N, approx.}$ Now

$$F = ma,$$

where a is the acceleration in m s^{-2} .

$$\therefore (T - 2) \times 10 = 2 \times a = 2 \times 0.2$$

$$\therefore T = 2.04 \text{ kgf} \quad (1)$$

(ii) When the lift descends with an acceleration of 10 cm s^{-2} or 0.1 m s^{-2} , the weight, 2 kgf, is now greater than T_1 kgf, the tension in the spring-balance.

$$\therefore \text{resultant force} = (2 - T_1) \text{ kgf} = (2 - T_1) \times 10 \text{ N approx.}$$

$$\therefore F = (2 - T_1) \times 10 = ma = 2 \times 0.1$$

$$\therefore T_1 = 2 - 0.02 = 1.98 \text{ kgf.}$$

(iii) When the lift moves with constant velocity, the acceleration is zero. In this case the reading on the spring-balance is exactly equal to the weight, 2 kgf.

Linear Momentum

Newton defined the force acting on an object as the rate of change of its momentum, the momentum being the product of its mass and velocity (p. 13). *Momentum is thus a vector quantity.* Suppose that the mass of an object is m , its initial velocity is u , and its final velocity due to a force F acting on it for a time t is v . Then

$$\text{change of momentum} = mv - mu,$$

and hence

$$F = \frac{mv - mu}{t}$$

$$\therefore Ft = mv - mu = \text{momentum change} \quad (1)$$

The quantity Ft (force \times time) is known as the *impulse* of the force on the object, and from (1) it follows that the units of momentum are the same as those of Pt , i.e., *newton second* (N s). From 'mass \times velocity', alternative units are 'kg m s⁻¹'.

Force and momentum change

A person of mass 50 kg who is jumping from a height of 5 metres will land on the ground with a velocity $= \sqrt{2gh} = \sqrt{2 \times 10 \times 5} = 10 \text{ m s}^{-1}$, assuming $g = 980 \text{ cm s}^{-2} = 10 \text{ m s}^{-2}$ approx. If he does not flex his knees on landing, he will be brought to rest very quickly, say in $\frac{1}{10}$ th second. The force F acting is then given by

$$F = \frac{\text{momentum change}}{\text{time}} \\ = \frac{50 \times 10}{\frac{1}{10}} = 5000 \text{ N} = 500 \text{ kgf (approx).}$$

This is a force of about 10 times the person's weight and this large force has a severe effect on the body.

Suppose, however, that the person flexes his knees and is brought to rest much more slowly on landing, say in 1 second. Then, from above, the force F now acting is 10 times less than before, or 50 kgf (approx). Consequently, much less damage is done to the person on landing.

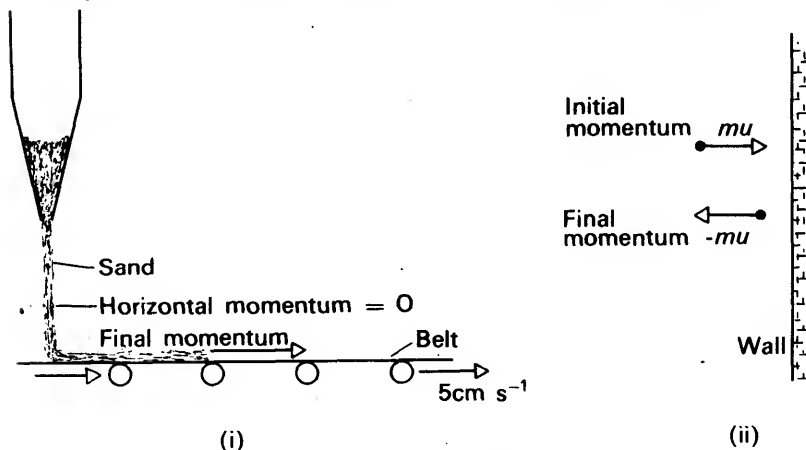


FIG. 1.13 Linear momentum

Suppose sand is allowed to fall vertically at a steady rate of 100 g s^{-1} on to a horizontal conveyor belt moving at a steady velocity of 5 cm s^{-1} . Fig. 1.13 (i). The initial horizontal velocity of the sand is zero. The final horizontal velocity is 5 cm s^{-1} . Now

$$\text{mass} = 100 \text{ g} = 0.1 \text{ kg}, \text{ velocity} = 5 \text{ cm s}^{-1} = 5 \times 10^{-2} \text{ m s}^{-1}$$

$$\therefore \text{momentum change per second} = 0.1 \times 5 \times 10^{-2} = 5 \times 10^{-3} \text{ newton} \\ = \text{force on belt}$$

Observe that this is a case where the *mass* changes with time and the velocity gained is constant. In terms of the calculus, the force is the rate of change of momentum mv , which is $v \times dm/dt$, and dm/dt is 100 g s^{-1} in this numerical example.

Consider a molecule of mass m in a gas, which strikes the wall of a vessel repeatedly with a velocity u and rebounds with a velocity $-u$. Fig. 1.13 (ii). Since momentum is a vector quantity, the momentum change = final momentum - initial momentum = $mu - (-mu) = 2mu$. If the containing vessel is a cube of side l , the molecule repeatedly takes a time $2l/u$ to make an impact with the same side.

$$\therefore \text{average force on wall due to molecule}$$

$$= \frac{\text{momentum change}}{\text{time}}$$

$$= \frac{2mu}{2l/u} = \frac{mu^2}{l}.$$

The total gas pressure is the average force per unit area on the walls of the container due to all the numerous gas molecules.

EXAMPLES

1. A hose ejects water at a speed of 20 cm s^{-1} through a hole of area 100 cm^2 . If the water strikes a wall normally, calculate the force on the wall in newton, assuming the velocity of the water normal to the wall is zero after collision.

$$\text{The volume of water per second striking the wall} = 100 \times 20 = 2000 \text{ cm}^3.$$

$$\therefore \text{mass per second striking wall} = 2000 \text{ g s}^{-1} = 2 \text{ kg s}^{-1}.$$

$$\text{Velocity change of water on striking wall} = 20 - 0 = 20 \text{ cm s}^{-1} = 0.2 \text{ m s}^{-1}.$$

$$\therefore \text{momentum change per second} = 2 (\text{kg s}^{-1}) \times 0.2 (\text{m s}^{-1}) = 0.4 \text{ newton}.$$

2. Sand drops vertically at the rate of 2 kg s^{-1} on to a conveyor belt moving horizontally with a velocity of 0.1 m s^{-1} . Calculate (i) the extra power needed to keep the belt moving, (ii) the rate of change of kinetic energy of the sand. Why is the power twice as great as the rate of change of kinetic energy?

$$\text{(i) Force required to keep belt moving} = \text{rate of increase of horizontal momentum of sand} = \text{mass per second} (dm/dt) \times \text{velocity change} = 2 \times 0.1 = 0.2 \text{ newton}.$$

$$\therefore \text{power} = \text{work done per second} = \text{force} \times \text{rate of displacement}$$

$$= \text{force} \times \text{velocity} = 0.2 \times 0.1 = 0.02 \text{ watt (p. 25)}.$$

(ii) Kinetic energy of sand = $\frac{1}{2}mv^2$.

\therefore rate of change of energy = $\frac{1}{2}v^2 \times \frac{dm}{dt}$, since v is constant,

$$= \frac{1}{2} \times 0.1^2 \times 2 = 0.01 \text{ watt.}$$

Thus the power supplied is twice as great as the rate of change of kinetic energy. The extra power is due to the fact that the sand does not immediately assume the velocity of the belt, so that the belt at first moves relative to the sand. The extra power is needed to overcome the friction between the sand and belt.

Conservation of Linear Momentum

We now consider what happens to the linear momentum of objects which *collide* with each other.

Experimentally, this can be investigated by several methods:

1. Trolleys in collision, with ticker-tapes attached to measure velocities.
2. Linear Air-track, using perspex models in collision and stroboscopic photography for measuring velocities.

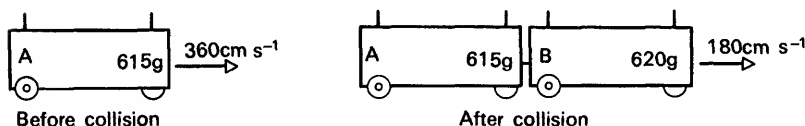


FIG. 1.14 Linear momentum experiment

As an illustration of the experimental results, the following measurements were taken in trolley collisions (Fig. 1.14):

Before collision.

Mass of trolley A = 615 g; initial velocity = 360 cm s^{-1} .

After collision.

A and B coalesced and both moved with velocity of 180 cm s^{-1} .

Thus the total linear momentum of A and B before collision = $0.615 \text{ (kg)} \times 3.6 \text{ (m s}^{-1}) + 0 = 2.20 \text{ kg m s}^{-1}$ (approx). The total momentum of A and B after collision = $1.235 \times 1.8 = 2.20 \text{ kg m s}^{-1}$ (approx).

Within the limits of experimental accuracy, it follows that *the total momentum of A and B before collision = the total momentum after collision*. Similar results are obtained if A and B are moving with different speeds after collision, or in opposite directions before collision.

Principle of Conservation of Linear Momentum

These experimental results can be shown to follow from Newton's second and third laws of motion (p. 12).

Suppose that a moving object A, of mass m_1 and velocity u_1 , collides

with another object B, of mass m_2 and velocity u_2 , moving in the same direction, Fig. 1.15. By Newton's law of action and reaction, the force F exerted by A on B is equal and opposite to that exerted by B on A. Moreover, the time t during which the force acted on B is equal to the time during which the force of reaction acted on A. Thus the magnitude of the impulse, Ft , on B is equal and opposite to the magnitude of the impulse on A. From equation (1), p. 16, the impulse is equal to the change of momentum. It therefore follows that the change in the total momentum of the two objects is *zero*, i.e., the total momentum of the two objects is constant although a collision had occurred. Thus if A moves with a reduced velocity v_1 after collision, and B then moves with an increased velocity v_2 ,

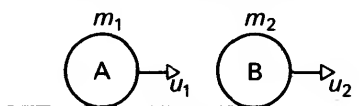


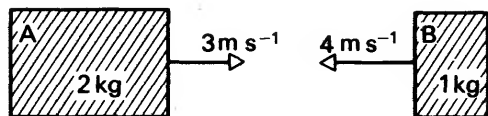
FIG. 1.15
Conservation of linear momentum

$$m_1 u_1 + m_2 u_2 = m_1 v_1 + m_2 v_2.$$

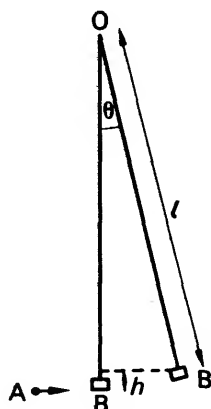
The principle of the conservation of linear momentum states that, if no external forces act on a system of colliding objects, the total momentum of the objects remains constant.

EXAMPLES

1. An object A of mass 2 kg is moving with a velocity of 3 m s^{-1} and collides head on with an object B of mass 1 kg moving in the opposite direction with a velocity of 4 m s^{-1} . Fig. 1.16 (i). After collision both objects coalesce, so that they move with a common velocity v . Calculate v .



(i)



(ii)

FIG. 1.16 Examples

Total momentum before collision of A and B in the direction of A

$$= 2 \times 3 - 1 \times 4 = 2 \text{ kg m s}^{-1}.$$

Note that momentum is a vector and the momentum of B is of opposite sign to A.

After collision, momentum of A and B in the direction of A = $2v + 1v = 3v$.

$$\therefore 3v = 2$$

$$\therefore v = \frac{2}{3} \text{ m s}^{-1}.$$

2. What is understood by (a) the principle of the *conservation of energy*, (b) the principle of the *conservation of momentum*?

A bullet of mass 20 g travelling horizontally at 100 m s^{-1} , embeds itself in the centre of a block of wood of mass 1 kg which is suspended by light vertical strings 1 m in length. Calculate the maximum inclination of the strings to the vertical.

Describe in detail how the experiment might be carried out and used to determine the velocity of the bullet just before the impact of the block. (N.)

Second part. Suppose A is the bullet, B is the block suspended from a point O, and θ is the maximum inclination to the vertical, Fig. 1.16(ii). If $v \text{ cm s}^{-1}$ is the common velocity of block and bullet when the latter is brought to rest relative to the block, then, from the principle of the conservation of momentum, since $20 \text{ g} = 0.02 \text{ kg}$,

$$(1 + 0.02)v = 0.02 \times 100$$

$$\therefore v = \frac{2}{1.02} = \frac{100}{51} \text{ m s}^{-1}.$$

The vertical height risen by block and bullet is given by $v^2 = 2gh$, where $g = 9.8 \text{ m s}^{-2}$ and $h = l - l \cos \theta = l(1 - \cos \theta)$.

$$\therefore v^2 = 2gl(1 - \cos \theta).$$

$$\therefore \left(\frac{100}{51}\right)^2 = 2 \times 9.8 \times 1(1 - \cos \theta).$$

$$\therefore 1 - \cos \theta = \left(\frac{100}{51}\right)^2 \times \frac{1}{2 \times 9.8} = 0.1962.$$

$$\therefore \cos \theta = 0.8038, \text{ or } \theta = 37^\circ \text{ (approx.)}$$

The velocity, v , of the bullet can be determined by applying the conservation of momentum principle.

Thus $mv = (m + M)V$, where m is the mass of the bullet, M is the mass of the block, and V is the common velocity. Then $v = (m + M)V/m$. The quantities m and M can be found by weighing. V is calculated from the horizontal displacement a of the block, since (i) $V^2 = 2gh$ and (ii) $h(2l - h) = a^2$ from the geometry of the circle, so that, to a good approximation, $2h = a^2/l$.

Inelastic and elastic collisions

In collisions, the total momentum of the colliding objects is always conserved. Usually, however, their total kinetic energy is not conserved. Some of it is changed to heat or sound energy, which is not recoverable. Such collisions are said to be *inelastic*. If the total kinetic energy is conserved, the collision is said to be *elastic*. The collision between two smooth billiard balls is approximately elastic. Many atomic collisions are elastic. Electrons may make elastic or inelastic collisions

with atoms of a gas. As proved on p. 28, the kinetic energy of a mass m moving with a velocity v has kinetic energy equal to $\frac{1}{2}mv^2$.

As an illustration of the mechanics associated with elastic collisions, consider a sphere A of mass m and velocity v incident on a stationary sphere B of equal mass m (Fig. 1.17 (i)). Suppose the collision is elastic, and after collision let A move with a velocity v_1 at an angle of 60° to its original direction and B move with a velocity v_2 at an angle θ to the direction of v .

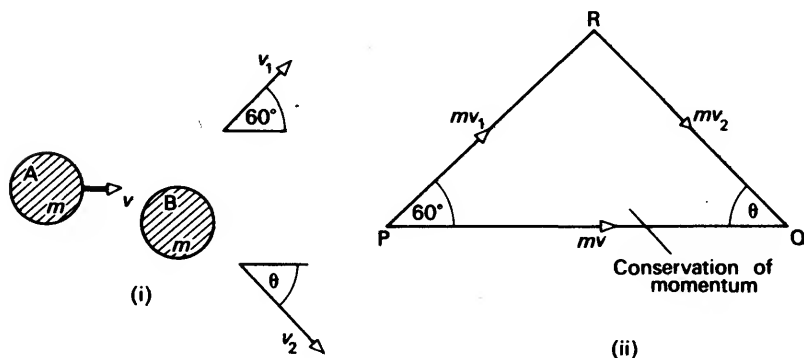


FIG. 1.17 Conservation of momentum

Since momentum is a vector (p. 17), we may represent the momentum mv of A by the line PQ drawn in the direction of v . Fig. 1.17 (ii). Likewise, PR represents the momentum mv_1 of A after collision. *Since momentum is conserved, the vector RQ must represent the momentum mv_2 of B after collision, that is,*

$$\vec{m}\vec{v} = \vec{m}\vec{v}_1 + \vec{m}\vec{v}_2.$$

Hence

$$\vec{v} = \vec{v}_1 + \vec{v}_2,$$

or PQ represents v in magnitude, PR represents v_1 and RQ represents v_2 . But if the collision is elastic,

$$\frac{1}{2}mv^2 = \frac{1}{2}mv_1^2 + \frac{1}{2}mv_2^2$$

$$\therefore v^2 = v_1^2 + v_2^2.$$

Consequently, triangle PRQ is a right-angled triangle with angle R equal to 90° .

$$\therefore v_1 = v \cos 60^\circ = \frac{v}{2}.$$

Also, $\theta = 90^\circ - 60^\circ = 30^\circ$, and $v_2 = v \cos 30^\circ = \frac{\sqrt{3}v}{2}$.

Coefficient of restitution

In practice, colliding objects do not stick together and kinetic energy is always lost. If a ball X moving with velocity u_1 collides head-on with a ball Y moving with a velocity u_2 in the same direction, then Y will move faster with a velocity v_1 say and X may then have a reduced velocity v_2 in the same direction. The coefficient of restitution, e , between X and Y is defined as the ratio:

$$\frac{\text{velocity of separation}}{\text{velocity of approach}} \quad \text{or} \quad \frac{v_2 - v_1}{u_1 - u_2}$$

The coefficient of restitution is approximately constant between two given materials. It varies from $e = 0$, when objects stick together and the collision is completely inelastic, to $e = 1$, when objects are very hard and the collision is practically elastic. Thus, from above, if $u_1 = 4 \text{ m s}^{-1}$, $u_2 = 1 \text{ m s}^{-1}$ and $e = 0.8$, then velocity of separation, $v_2 - v_1 = 0.8 \times (4 - 1) = 2.4 \text{ m s}^{-1}$.

Momentum and Explosive forces

There are numerous cases where momentum changes are produced by *explosive* forces. An example is a bullet of mass $m = 50 \text{ g}$ say, fired from a rifle of mass $M = 2 \text{ kg}$ with a velocity v of 100 m s^{-1} . Initially, the total momentum of the bullet and rifle is zero. From the principle of the conservation of linear momentum, when the bullet is fired the total momentum of bullet and rifle is still zero, since no external force has acted on them. Thus if V is the velocity of the rifle,

$$mv (\text{bullet}) + MV (\text{rifle}) = 0$$

$$\therefore MV = -mv, \quad \text{or} \quad V = -\frac{m}{M}v.$$

The momentum of the rifle is thus *equal and opposite* to that of the bullet. Further, $V/v = -m/M$. Since $m/M = 50/2000 = 1/40$, it follows that $V = -v/40 = 2.5 \text{ m s}^{-1}$. This means that the rifle moves back or *recoils* with a velocity only about $\frac{1}{40}$ th that of the bullet.

If it is preferred, one may also say that the explosive force produces the same numerical momentum change in the bullet as in the rifle. Thus $mv = MV$, where V is the velocity of the rifle in the *opposite* direction to that of the bullet. The joule (J) is the unit of energy (p. 24).

The kinetic energy, E_1 , of the bullet $= \frac{1}{2}mv^2 = \frac{1}{2} \cdot 0.05 \cdot 100^2 = 250 \text{ J}$

The kinetic energy, E_2 , of the rifle $= \frac{1}{2}MV^2 = \frac{1}{2} \cdot 2 \cdot 2.5^2 = 6.25 \text{ J}$

Thus the total kinetic energy produced by the explosion $= 256.25 \text{ J}$. The kinetic energy E_1 of the bullet is thus $250/256.25$, or about 98%, of the total energy. This is explained by the fact that the kinetic energy depends on the *square* of the velocity. The high velocity of the bullet thus more than compensates for its small mass relative to that of the rifle. See also p. 26.

Rocket

Consider a rocket moving in outer space where no external forces act on it. Suppose its mass is M and its velocity is v at a particular instant. Fig. 1.18 (i). When a mass m of fuel is ejected, the mass of the rocket becomes $(M - m)$ and its velocity increases to $(v + \Delta v)$. Fig. 1.18 (ii).

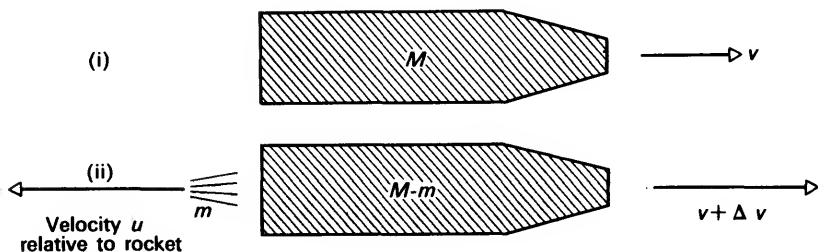


FIG. 1.18. Motion of rocket

Suppose the fuel is always ejected at a constant speed u relative to the rocket. Then the velocity of the mass $m = v + \frac{\Delta v}{2} - u$ in the direction of the rocket, since the initial velocity of the rocket is v and the final velocity is $v + \Delta v$, an average of $v + \Delta v/2$.

We now apply the principle of the conservation of momentum to the rocket and fuel. *Initially*, before m of fuel was ejected, momentum of rocket and fuel inside rocket $= Mv$.

After m is ejected, momentum of rocket $= (M-m)(v + \Delta v)$

$$\text{and momentum of fuel} = m \left(v + \frac{\Delta v}{2} - u \right).$$

$$\therefore (M-m)(v + \Delta v) + m \left(v + \frac{\Delta v}{2} - u \right) = Mv.$$

Neglecting the product of $m \cdot \Delta v$, then, after simplification,

$$M \cdot \Delta v - mu = 0,$$

$$\therefore \frac{m}{M} = \frac{\Delta v}{u}.$$

Now

$$m = \text{mass of fuel ejected} = -\Delta M,$$

$$\therefore -\frac{\Delta M}{M} = \frac{\Delta v}{u}.$$

Integrating between limits of M, M_0 and v, v_0 respectively

$$\therefore \int_{M_0}^M -\frac{\Delta M}{M} = \frac{1}{u} \int_{v_0}^v \Delta v.$$

$$\therefore -\log_e \frac{M}{M_0} = \frac{v - v_0}{u}.$$

$$\therefore M = M_0 e^{-(v - v_0)/u} \quad (1)$$

or

$$v = v_0 - u \log_e (M/M_0) \quad (2)$$

When the mass M decreases to $M_0/2$

$$v = v_0 + u \log_e 2.$$

Motion of centre of mass

If two particles, masses m_1 and m_2 , are distances x_1, x_2 respectively from a given axis, their *centre of mass* is at a distance \bar{x} from the axis given by $m_1x_1 + m_2x_2 = (m_1 + m_2)\bar{x}$. See p. 104. Since velocity, $v = dx/dt$ generally, the velocity \bar{v} of the centre of mass in the particular direction is given by $m_1v_1 + m_2v_2 = (m_1 + m_2)\bar{v}$, where v_1, v_2 are the respective velocities of m_1, m_2 . The quantity $(m_1v_1 + m_2v_2)$ represents the total momentum of the two particles. The quantity $(m_1 + m_2)\bar{v} = M\bar{v}$, where M is the total mass of the particles. Thus we can imagine that the total mass of the particles is concentrated at the centre of mass while they move, and that the velocity \bar{v} of the centre of mass is always given by *total momentum* $= M\bar{v}$.

If *internal forces* act on the particles while moving, then, since action and reaction are equal and opposite, their resultant on the whole body is zero. Consequently the total momentum is unchanged and hence the velocity or motion of their centre of mass is unaffected. If an *external force*, however, acts on the particles, the total momentum is changed. The motion of their centre of mass now follows a path which is due to the external force.

We can apply this to the case of a shell fired from a gun. The centre of mass of the shell follows at first a parabolic path. This is due to the external force of gravity, its weight. If the shell explodes in mid-air, the fragments fly off in different directions. But the numerous internal forces which occur in the explosion have zero resultant, since action and reaction are equal and opposite and the forces can all be paired. Consequently *the centre of mass of all the fragments continues to follow the same parabolic path*. As soon as one fragment reaches the ground, an external force now acts on the system of particles. A different parabolic path is then followed by the centre of mass.

If a bullet is fired in a horizontal direction from a rifle, where is their centre of mass while the bullet and rifle are both moving?

Work

When an engine pulls a train with a constant force of 50 units through a distance of 20 units in its own direction, the engine is said by definition to do an amount of *work* equal to 50×20 or 1000 units, the product of the force and the distance. Thus if W is the amount of work,

$$W = \text{force} \times \text{distance moved in direction of force.}$$

Work is a *scalar* quantity; it has no property of direction but only magnitude. When the force is one newton and the distance moved is one metre, then the work done is one *joule*. Thus a force of 50 N moving through a distance of 10 m does 50×10 or 500 joule of work. Note this is also a measure of the *energy* transferred to the object.

The force to raise steadily a mass of 1 kg is 1 kilogram force (1 kgf), which is about 10 N (see p. 14). Thus if the mass of 1 kg is raised vertically through 1 m, then, approximately, work done $= 10(\text{N}) \times 1(\text{m}) = 10 \text{ joule}$.

The c.g.s. unit of work is the *erg*; it is the work done when a force of 1 dyne moves through 1 cm. Since $1 \text{ N} = 10^5$ dynes and $1 \text{ m} = 100 \text{ cm}$, then 1 N moving through 1 m does an amount of work = 10^5 (dyne) \times 100 (cm) = 10^7 ergs = 1 joule, by definition of the joule (p. 24).

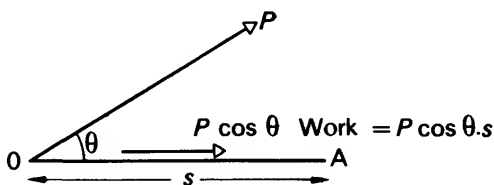


FIG. 1.19 Work

Before leaving the topic of 'work', the reader should note carefully that we have assumed the force to move an object in its own direction. Suppose, however, that a force P pulls an object a distance s along a line OA acting at an angle θ to it, Fig. 1.19. The component of P along OA is $P \cos \theta$ (p. 8), and this is the effective part of P pulling along the direction OA . The component of P along a direction perpendicular to OA has no effect along OA . Consequently

$$\text{work done} = P \cos \theta \times s.$$

In general, the work done by a force is equal to the product of the force and the displacement of its point of application in the direction of the force.

Power

When an engine does work quickly, it is said to be operating at a high *power*; if it does work slowly it is said to be operating at a low power. 'Power' is defined as the *work done per second*, i.e.,

$$\text{power} = \frac{\text{work done}}{\text{time taken}}.$$

The practical unit of power, the SI unit, is 'joule per second' or *watt* (W); the watt is defined as the rate of working at 1 joule per second.

$$1 \text{ horse-power (hp)} = 746 \text{ W} = \frac{3}{4} \text{ kW (approx),}$$

where $1 \text{ kW} = 1$ kilowatt of 1000 watt. Thus a small motor of $\frac{1}{6}$ hp in a vacuum carpet cleaner has a power of about 125 W.

Kinetic Energy

An object is said to possess *energy* if it can do work. When an object possesses energy because it is moving, the energy is said to be *kinetic*, e.g., a flying stone can disrupt a window. Suppose that an object of mass m is moving with a velocity u , and is gradually brought to rest in a distance s by a constant force F acting against it. The kinetic energy originally possessed by the object is equal to the work done against F , and hence

$$\text{kinetic energy} = F \times s.$$

But $F = ma$, where a is the retardation of the object. Hence $F \times s = mas$. From $v^2 = u^2 + 2as$ (see p. 6), we have, since $v = 0$ and a is negative in this case,

$$0 = u^2 - 2as, \text{ i.e., } as = \frac{u^2}{2}.$$

$$\therefore \text{kinetic energy} = mas = \frac{1}{2}mu^2.$$

When m is in kg and u is in m s^{-1} , then $\frac{1}{2}mu^2$ is in *joule*. Thus a car of mass 1000 kg, moving with a velocity of 36 km h^{-1} or 10 m s^{-1} , has an amount W of kinetic energy given by

$$W = \frac{1}{2}mu^2 = \frac{1}{2} \times 1000 \times 10^2 = 50000 \text{ J}$$

Kinetic Energies due to Explosive Forces

Suppose that, due to an explosion or nuclear reaction, a particle of mass m breaks away from the total mass concerned and moves with velocity v , and a mass M is left which moves with velocity V in the opposite direction. Then

$$\frac{\text{kinetic energy, } E_1, \text{ of mass } m}{\text{kinetic energy, } E_2, \text{ of mass } M} = \frac{\frac{1}{2}mv^2}{\frac{1}{2}MV^2} = \frac{mv^2}{MV^2} \quad (1)$$

Now from the principle of the conservation of linear momentum, $mv = MV$. Thus $v = MV/m$. Substituting for v in (1).

$$\therefore \frac{E_1}{E_2} = \frac{mM^2V^2}{m^2MV^2} = \frac{M}{m} = \frac{1/m}{1/M}$$

Hence the energy is *inversely*-proportional to the masses of the particles, that is, the smaller mass, m say, has the larger energy. Thus if E is the total energy of the two masses, the energy of the smaller mass $= ME/(M+m)$. An α -particle has a mass of 4 units and a radium nucleus a mass of 228 units. If disintegration of a thorium nucleus, mass 232, produces an α -particle and radium nucleus, and a release of energy of 4.05 MeV, where $1 \text{ MeV} = 1.6 \times 10^{-13} \text{ J}$, then

$$\text{energy of } \alpha\text{-particle} = \frac{228}{(4+228)} \times 4.05 = 3.98 \text{ MeV}.$$

The α -particle thus travels a relatively long distance before coming to rest compared to the radium nucleus.

Potential Energy

A weight held stationary above the ground has energy, because, when released, it can raise another object attached to it by a rope passing over a pulley, for example. A coiled spring also has energy, which is released gradually as the spring uncoils. The energy of the weight or spring is called *potential energy*, because it arises from the position or arrangement of the body and not from its motion. In the case of the

weight, the energy given to it is equal to the work done by the person or machine which raises it steadily to that position against the force of attraction of the earth. In the case of the spring, the energy is equal to the work done in displacing the molecules from their normal equilibrium positions against the forces of attraction of the surrounding molecules.

If the mass of an object is m , and the object is held stationary at a height h above the ground, the energy released when the object falls to the ground is equal to the work done

$$= \text{force} \times \text{distance} = \text{weight of object} \times h.$$

Suppose the weight is 5 kgf and h is 4 metre. Then, since 1 kgf = 9.8 N = 10 N approx, then

$$\text{potential energy P.E.} = 50 \text{ (N)} \times 4 \text{ (m)} = 200 \text{ J}$$

$$(\text{more accurately, P.E.} = 192 \text{ J}).$$

Generally, at a height of h ,

$$\text{potential energy} = mgh,$$

where m is in kg, h is in metre, $g = 9.8$.

EXAMPLE

Define *work*, *kinetic energy*, *potential energy*. Give *one* example of *each* of the following: (a) the conversion into kinetic energy of the work done on a body and (b) the conversion into potential energy of the work done on a body.

A rectangular block of mass 10 g rests on a rough plane which is inclined to the horizontal at an angle $\sin^{-1}(0.05)$. A force of 0.03 newton, acting in a direction parallel to a line of greatest slope, is applied to the block so that it moves up the plane. When the block has travelled a distance of 110 cm from its initial position, the applied force is removed. The block moves on and comes to rest again after travelling a further 25 cm. Calculate (i) the work done by the applied force, (ii) the gain in potential energy of the block and (iii) the value of the coefficient of sliding friction between the block and the surface of the inclined plane. How would the coefficient of sliding friction be measured if the angle of the slope could be altered? (O. and C.)

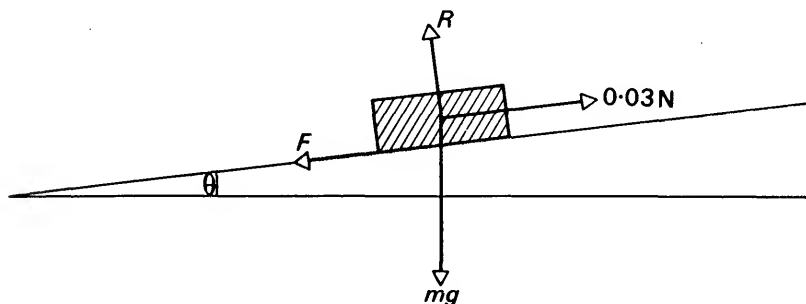


FIG. 1.20 Example

(i) Force = 0.03 newton; distance = 110 cm = 1.1 m.

$$\therefore \text{work} = 0.03 \times 1.1 = 0.033 \text{ J}.$$

$$\begin{aligned} \text{(ii) Gain in P.E.} &= wt \times \text{height moved} = 0.01 \text{ kgf} \times 1.35 \sin \theta \text{ m,} \\ &= 0.01 \times 9.8 \text{ newton} \times 1.35 \times 0.05 \text{ m} = 0.0066 \text{ J (approx.).} \end{aligned}$$

$$\begin{aligned} \text{(iii) Work done against frictional force } F &= \text{work done by force} - \text{gain in P.E.} \\ &= 0.033 - 0.0066 = 0.0264 \text{ J.} \end{aligned}$$

$$\therefore F \times 1.35 = 0.0264.$$

$$\therefore F = \frac{0.0264}{1.35} \text{ newton.}$$

$$\text{Normal reaction, } R = mg \cos \theta = mg \text{ (approx.), since } \theta \text{ is so small}$$

$$\therefore \mu = \frac{F}{R} = \frac{0.0264}{1.35 \times 0.01 \times 9.8} = 0.2 \text{ (approx.).}$$

Conservative Forces

If a ball of weight W is raised steadily from the ground to a point X at a height h above the ground, the work done is $W.h$. The potential energy, P.E., of the ball is thus $W.h$. Now whatever route is taken from ground level to X , the work done is the same—if a *longer* path is chosen, for example, the component of the weight in the particular direction must then be overcome and so the force required to move the ball is correspondingly smaller. The P.E. of the ball at X is thus independent of the route to X . This implies that if the ball is taken in a closed path round to X again, *the total work done is zero*. Work has been expended on one part of the closed path, and regained on the remaining part.

When the work done in moving round a closed path in a field to the original point is zero, the forces in the field are called *conservative forces*. The earth's gravitational field is an example of a field containing conservative forces, as we now show.

Suppose the ball falls from a place Y at a height h to another X at a height of x above the ground. Fig. 1.21. Then, if W is the weight of the ball and m its mass,

$$\text{P.E. at } X = Wx = mgx$$

$$\text{and K.E. at } X = \frac{1}{2}mv^2 = \frac{1}{2}m \cdot 2g(h-x) = mg(h-x),$$

using $v^2 = 2as = 2g(h-x)$. Hence

$$\text{P.E.} + \text{K.E.} = mgx + mg(h-x) = mgh.$$

Thus at any point such as X , the total mechanical energy of the falling ball is equal to the original energy. The mechanical energy is hence constant or conserved. This is the case for a conservative field.

Non-Conservative forces. Principle of Conservation of Energy

The work done in taking a mass m round a closed path in the conservative earth's gravitational field is zero. Fig. 1.22 (i). If the work done in taking an object round a closed path to its original position is

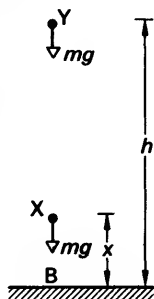


FIG. 1.21.
Mechanical energy

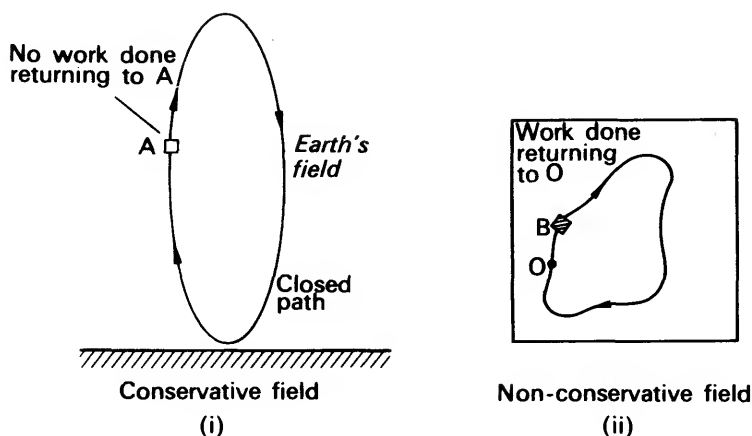


FIG. 1.22 Non-conservative and conservative fields

not zero, the forces in the field are said to be *non-conservative*. This is the case, for example, when a wooden block B is pushed round a closed path on a rough table to its initial position O. Work is therefore done against friction, both as A moves away from O and as it returns. In a conservative field, however, work is done during part of the path and regained for the remaining part.

When a body falls in the earth's gravitational field, a small part of the energy is used up in overcoming the resistance of the air. This energy is dissipated or lost as heat—it is not regained in moving the body back to its original position. This resistance is another example of the action of a non-conservative force.

Although energy may be transformed from one form to another, as in the last example from mechanical energy to heat, *the total energy in a given system is always constant*. If an electric motor is supplied with 1000 joule of energy, 850 joule of mechanical energy, 140 joule of heat energy and 10 joule of sound energy may be produced. This is called the *Principle of the Conservation of Energy* and is one of the key principles in science.

Mass and Energy

Newton said that the 'mass' of an object was 'a measure of the quantity of matter' in it. In 1905, Einstein showed from his Special Theory of Relativity that energy is released from an object when its mass decreases. His mass-energy relation states that if the mass decreases by Δm kg, the energy released in joule, ΔW , is given by

$$\Delta W = \Delta m \cdot c^2,$$

where c is the numerical value of the speed of light in m s^{-1} , which is 3×10^8 . Experiments in Radioactivity on nuclear reactions showed that Einstein's relation was true. Thus mass is a form of energy.

Einstein's relation shows that even if a small change in mass occurs, a

relatively large amount of energy is produced. Thus if $\Delta m = 1$ milligramme $= 10^{-6}$ kg, the energy ΔW released

$$= \Delta m \cdot c^2 = 10^{-6} \times (3 \times 10^8)^2 = 9 \times 10^{10} \text{ J.}$$

This energy will keep 250 000 100-W lamps burning for about an hour. In practice, significant mass changes occur only in nuclear reactions.

The internal energy of a body of mass m may be considered as $E_{\text{int}} = mc^2$, where m is its rest mass. In nuclear reactions where two particles collide, a change occurs in their total kinetic energy and in their total mass. The increase in total kinetic energy is accompanied by an equal decrease in internal energy, $\Delta m \cdot c^2$. Thus the total energy, kinetic plus internal, remains constant.

Before Einstein's mass-energy relation was known, two independent laws of science were:

(1) *The Principle of the Conservation of Mass* (the total mass of a given system of objects is constant even though collisions or other actions took place between them);

(2) *The Principle of the Conservation of Energy* (the total energy of a given system is constant). From Einstein's relation, however, the two laws can be combined into one, namely, the Principle of the Conservation of Energy.

The summary below may assist the reader; it refers to the units of some of the quantities encountered, and their relations.

Quantity	SI	C.G.S.	Relations
Force (vector)	newton (N)	dyne	$10^5 \text{ dyne} = 1 \text{ N}$ $1 \text{ kgf} = 9.8 \text{ N}$ (approx, 10 N) $1 \text{ gf} = 0.0098 \text{ N}$ (approx, 0.01 N)
Mass (scalar)	kilogramme (kg)	gramme (g)	$1000 \text{ g} = 1 \text{ kg}$
Momentum (vector)	newtonsecond (Ns)	dyne second	$10^5 \text{ dyn s} = 1 \text{ N s}$
Energy (scalar)	joule (J)	erg	$10^7 \text{ erg} = 1 \text{ J}$
Power (scalar)	watt (W)	erg s^{-1}	$1 \text{ W} = 1 \text{ J s}^{-1}$ $1 \text{ h.p.} = 746 \text{ W}$

Dimensions

By the *dimensions* of a physical quantity we mean the way it is related to the fundamental quantities mass, length and time; these are usually denoted by M, L, and T respectively. An area, length \times breadth, has dimensions $L \times L$ or L^2 ; a volume has dimensions L^3 ; density, which is mass/volume, has dimensions M/L^3 or ML^{-3} ; relative density has no dimensions, since it is the ratio of similar quantities, in this case two masses (p. 114); an angle has no dimensions, since it is the ratio of two lengths.

As an area has dimensions L^2 , the *unit* may be written in terms of the metre as 'm²'. Similarly, the dimensions of a volume are L^3 and hence

the unit is 'm³'. Density has dimensions ML^{-3} . The density of mercury is thus written as '13600 kg m⁻³'. If some physical quantity has dimensions $ML^{-1}T^{-1}$, its unit may be written as 'kg m⁻¹ s⁻¹'.

The following are the dimensions of some quantities in Mechanics:

Velocity. Since velocity = $\frac{\text{distance}}{\text{time}}$, its dimensions are L/T or LT^{-1} .

Acceleration. The dimensions are those of velocity/time, i.e., L/T^2 or LT^{-2} .

Force. Since force = mass \times acceleration, its dimensions are MLT^{-2} .

Work or Energy. Since work = force \times distance, its dimensions are ML^2T^{-2} .

EXAMPLE

In the gas equation $(p + \frac{a}{V^2})(V - b) = RT$, what are the dimensions of the constants a and b ?

p represents pressure, V represents volume. The quantity a/V^2 must represent a pressure since it is added to p . The dimensions of $p = [\text{force}]/[\text{area}] = MLT^{-2}/L^2 = ML^{-1}T^{-2}$; the dimensions of $V = L^3$. Hence

$$\frac{[a]}{L^6} = ML^{-1}T^{-2}, \text{ or } [a] = ML^5T^{-2}.$$

The constant b must represent a volume since it is subtracted from V . Hence

$$[b] = L^3.$$

Application of Dimensions. Simple Pendulum

If a small mass is suspended from a long thread so as to form a simple pendulum, we may reasonably suppose that the period, T , of the oscillations depends only on the mass m , the length l of the thread, and the acceleration, g , due to gravity at the place concerned. Suppose then that

$$T = km^x l^y g^z \quad \dots \quad (i)$$

where x, y, z, k are unknown numbers. The dimensions of g are LT^{-2} from above. Now the dimensions of both sides of (i) must be the same.

$$\therefore T = M^x L^y (LT^{-2})^z.$$

Equating the indices of M, L, T on both sides, we have

$$x = 0,$$

$$y + z = 0,$$

and

$$-2z = 1.$$

$$\therefore z = -\frac{1}{2}, y = \frac{1}{2}, x = 0.$$

Thus, from (i), the period T is given by

$$T = k l^{\frac{1}{2}} g^{-\frac{1}{2}},$$

or

$$T = k \sqrt{\frac{l}{g}}.$$

We cannot find the magnitude of k by the method of dimensions, since it is a number. A complete mathematical investigation shows that $k = 2\pi$ in this case, and hence $T = 2\pi\sqrt{l/g}$. (See also p. 48).

Velocity of Transverse Wave in a String

As another illustration of the use of dimensions, consider a wave set up in a stretched string by plucking it. The velocity, V , of the wave depends on the tension, F , in the string, its length l , and its mass m , and we can therefore suppose that

$$V = kF^x l^y m^z, \quad (i)$$

where x, y, z are numbers we hope to find by dimensions and k is a constant.

The dimensions of velocity, V , are LT^{-1} , the dimensions of tension, F , are MLT^{-2} , the dimension of length, l , is L , and the dimension of mass, m , is M . From (i), it follows that

$$LT^{-1} \equiv (MLT^{-2})^x \times L^y \times M^z.$$

Equating powers of M, L , and T on both sides,

$$\therefore 0 = x + z, \quad (i)$$

$$1 = x + y, \quad (ii)$$

and $-1 = -2x, \quad (iii)$

$$\therefore x = \frac{1}{2}, z = -\frac{1}{2}, y = \frac{1}{2}.$$

$$\therefore V = k \cdot F^{\frac{1}{2}} l^{\frac{1}{2}} m^{-\frac{1}{2}},$$

$$\text{or } V = k \sqrt{\frac{Fl}{m}} = k \sqrt{\frac{F}{m/l}} = k \sqrt{\frac{\text{Tension}}{\text{mass per unit length}}}$$

A complete mathematical investigation shows that $k = 1$.

The method of dimensions can thus be used to find the relation between quantities when the mathematics is too difficult. It has been extensively used in hydrodynamics, for example. See also pp. 176, 181.

EXERCISES 1

(Assume $g = 10 \text{ m s}^{-2}$, unless otherwise given)

What are the missing words in the statements 1–10?

1. The dimensions of velocity are ...
2. The dimensions of force are ...
3. Using 'vector' or 'scalar', (i) mass is a ... (ii) force is a ... (iii) energy is a ... (iv) momentum is a ...
4. Linear momentum is defined as ...
5. An 'elastic' collision is one in which the ... and the ... are conserved.
6. When two objects collide, their ... is constant provided no ... forces act.
7. One newton \times one metre = ...

8. 1 kilogram force = ... newton, approx.
9. The momentum of two different bodies must be added by a ... method.
10. Force is the ... of change of momentum.

Which of the following answers, A, B, C, D or E, do you consider is the correct one in the statements 11–14?

11. When water from a hosepipe is incident horizontally on a wall, the force on the wall is calculated from *A* speed of water, *B* mass \times velocity, *C* mass per second \times velocity, *D* energy of water, *E* momentum change.

12. When a ball of mass 2 kg moving with a velocity of 10 m s^{-1} collides head-on with a ball of mass 3 kg and both move together after collision, the common velocity is *A* 5 m s^{-1} and energy is lost, *B* 4 m s^{-1} and energy is lost, *C* 2 m s^{-1} and energy is gained, *D* 6 m s^{-1} and momentum is gained, *E* 6 m s^{-1} and energy is conserved.

13. An object of mass 2 kg moving with a velocity of 4 m s^{-1} has a kinetic energy of *A* 8 joule, *B* 16 erg, *C* 4000 erg, *D* 16 joule, *E* 40000 joule.

14. The dimensions of work are *A* ML^2T^{-2} and it is a scalar, *B* ML^2T^{-2} and it is a vector, *C* MLT^{-1} and it is a scalar, *D* ML^2T and it is a scalar, *E* MLT and it is a vector.

15. A car moving with a velocity of 36 km h^{-1} accelerates uniformly at 1 m s^{-2} until it reaches a velocity of 54 km h^{-1} . Calculate (i) the time taken, (ii) the distance travelled during the acceleration, (iii) the velocity reached 100 m from the place where the acceleration began.

16. A ball of mass 100 g is thrown vertically upwards with an initial speed of 72 km h^{-1} . Calculate (i) the time taken to return to the thrower, (ii) the maximum height reached, (iii) the kinetic and potential energies of the ball half-way up.

17. The velocity of a ship A relative to a ship B is 10.0 km h^{-1} in a direction N. 45° E. If the velocity of B is 20.0 km h^{-1} in a direction N. 60° W., find the actual velocity of A in magnitude and direction.

18. Calculate the energy of (i) a 2 kg object moving with a velocity of 10 m s^{-1} , (ii) a 10 kg object held stationary 5 m above the ground.

19. A 4 kg ball moving with a velocity of 10.0 m s^{-1} collides with a 16 kg ball moving with a velocity of 4.0 m s^{-1} (i) in the same direction, (ii) in the opposite direction. Calculate the velocity of the balls in each case if they coalesce on impact, and the loss of energy resulting from the impact. State the principle used to calculate the velocity.

20. A ship X moves due north at 30.0 km h^{-1} ; a ship Y moves N. 60° W. at 20.0 km h^{-1} . Find the velocity of Y relative to X in magnitude and direction. If Y is 10 km due east of X at this instant, find the closest distance of approach of the two ships.

21. Two buckets of mass 6 kg are each attached to one end of a long inextensible string passing over a fixed pulley. If a 2 kg mass of putty is dropped from a height of 5 m into one bucket, calculate (i) the initial velocity of the system, (ii) the acceleration of the system, (iii) the loss of energy of the 2 kg mass due to the impact.

22. A bullet of mass 25 g and travelling horizontally at a speed of 200 m s^{-1} imbeds itself in a wooden block of mass 5 kg suspended by cords 3 m long. How far will the block swing from its position of rest before beginning to return? Describe a suitable method of suspending the block for this experiment and explain briefly the principles used in the solution of the problem. (L.)

23. State the principle of the conservation of linear momentum and show how it follows from Newton's laws of motion.

A stationary radioactive nucleus of mass 210 units disintegrates into an alpha particle of mass 4 units and a residual nucleus of mass 206 units. If the kinetic energy of the alpha particle is E , calculate the kinetic energy of the residual nucleus. (N.)

24. Define linear momentum and state the principle of conservation of linear momentum. Explain briefly how you would attempt to verify this principle by experiment.

Sand is deposited at a uniform rate of 20 kilogramme per second and with negligible kinetic energy on to an empty conveyor belt moving horizontally at a constant speed of 10 metre per minute. Find (a) the force required to maintain constant velocity, (b) the power required to maintain constant velocity, and (c) the rate of change of kinetic energy of the moving sand. Why are the latter two quantities unequal? (O. & C.)

25. What do you understand by the *conservation of energy*? Illustrate your answer by reference to the energy changes occurring (a) in a body whilst falling to and on reaching the ground, (b) in an X-ray tube.

The constant force resisting the motion of a car, of mass 1500 kg, is equal to one-fifteenth of its weight. If, when travelling at 48 km per hour, the car is brought to rest in a distance of 50 m by applying the brakes, find the additional retarding force due to the brakes (assumed constant) and the heat developed in the brakes. (N.)

26. Define *uniform acceleration*. State, for each case, one set of conditions sufficient for a body to describe (a) a parabola, (b) a circle.

A projectile is fired from ground level, with velocity 500 m s^{-1} at 30° to the horizontal. Find its horizontal range, the greatest vertical height to which it rises, and the time to reach the greatest height. What is the least speed with which it could be projected in order to achieve the same horizontal range? (The resistance of the air to the motion of the projectile may be neglected.) (O.)

27. Define *momentum* and state the *law of conservation of linear momentum*.

Discuss the conservation of linear momentum in the following cases (a) a freely falling body strikes the ground without rebounding, (b) during free flight an explosive charge separates an earth satellite from its propulsion unit, (c) a billiard ball bounces off the perfectly elastic cushion of a billiard table.

A bullet of mass 10 g travelling horizontally with a velocity of 300 m s^{-1} strikes a block of wood of mass 290 g which rests on a rough horizontal floor. After impact the block and bullet move together and come to rest when the block has travelled a distance of 15 m. Calculate the coefficient of sliding friction between the block and the floor. (O. & C.)

28. Explain the distinction between *fundamental* and *derived* units, using two examples of each.

Derive the dimensions of (a) the *moment of a couple* and *work*, and comment on the results, (b) the constants a and b in van der Waals' equation $(p + a/v^2)(v - b) = rT$ for unit mass of a gas. (N.)

29. Explain what is meant by the relative velocity of one moving object with respect to another.

A ship A is moving eastward with a speed of 15 km h^{-1} and another ship B , at a given instant 10 km east of A , is moving southwards with a speed of 20 km h^{-1} . How long after this instant will the ships be nearest to each other, how far apart will they be then, and in what direction will B be sighted from A ? (C.)

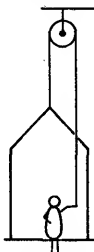
30. Define *momentum* and state the *law of conservation of linear momentum*.

Outline an experiment to demonstrate momentum conservation and discuss the accuracy which could be achieved.

Show that in a collision between two moving bodies in which no external act, the conservation of linear momentum may be deduced directly from Newton's laws of motion.

A small spherical body slides with velocity v and without rolling on a smooth horizontal table and collides with an identical sphere which is initially at rest on the table. After the collision the two spheres slide without rolling away from the point of impact, the velocity of the first sphere being in a direction at 30° to its previous velocity. Assuming that energy is conserved, and that there are no horizontal external forces acting, calculate the speed and direction of travel of the target sphere away from the point of impact. (O. & C.)

31. Answer the following questions making particular reference to the physical principles concerned (a) explained why the load on the back wheels of a motor car increases when the vehicle is accelerating, (b) the diagram, Fig. 1.23, shows a painter in a crate which hangs alongside a building. When the painter who weighs 100 kgf pulls on the rope the force he exerts on the floor of the crate is 45 kgf . If the crate weighs 25 kgf find the acceleration. (N.)



32. Derive an expression for the kinetic energy of a moving body. Fig. 1.23

A vehicle of mass 2000 kg travelling at 10 m s^{-1} on a horizontal surface is brought to rest in a distance of 12.5 m by the action of its brakes. Calculate the average retarding force. What horse-power must the engine develop in order to take the vehicle up an incline of 1 in 10 at a constant speed of 10 m s^{-1} if the frictional resistance is equal to 20 kgf ? (L.)

33. Explain what is meant by the principle of conservation of energy for a system of particles not acted upon by any external forces. What modifications are introduced when external forces are operative?

A bobsleigh is travelling at 10 m s^{-1} when it starts ascending an incline of 1 in 100 . If it comes to rest after travelling 150 m up the slope, calculate the proportion of the energy lost in friction and deduce the coefficient of friction between the runners and the snow. (O. & C.)

34. State Newton's Laws of Motion and deduce from them the relation between the distance travelled and the time for the case of a body acted upon by a constant force. Explain the units in which the various quantities are measured.

A fire engine pumps water at such a rate that the velocity of the water leaving the nozzle is 15 m s^{-1} . If the jet be directed perpendicularly on to a wall and the rebound of the water be neglected, calculate the pressure on the wall (1 m^3 water weighs 1000 kg). (O. & C.)

chapter two

Circular motion. S.H.M. Gravitation

Angular Velocity

IN the previous chapter we discussed the motion of an object moving in a straight line. There are numerous cases of objects moving in a

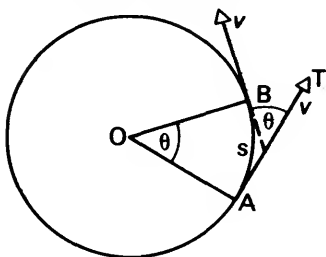


FIG. 2.1 Circular motion

curve about some fixed point. The earth and the moon revolve continuously round the sun, for example, and the rim of the balance-wheel of a watch moves to-and-fro in a circular path about the fixed axis of the wheel. In this chapter we shall study the motion of an object moving in a circle with a *uniform speed* round a fixed point O as centre, Fig. 2.1.

If the object moves from A to B so that the radius OA moves through an angle θ , its *angular velocity*, ω , about O is defined as the *change of the angle per second*. Thus if t is the time taken by the object to move from A to B,

$$\omega = \frac{\theta}{t} \quad (1)$$

Angular velocity is usually expressed in 'radian per second' (rad s^{-1}). From (1),

$$\theta = \omega t \quad (2)$$

which is analogous to the formula 'distance = uniform velocity \times time' for motion in a straight line. It will be noted that the time T to describe the circle once, known as the *period* of the motion, is given by

$$T = \frac{2\pi}{\omega} \quad (3)$$

since 2π radians = 360° by definition.

If s is the length of the arc AB, then $s/r = \theta$, by definition of an angle in radians.

$$\therefore s = r\theta.$$

Dividing by t , the time taken to move from A to B,

$$\therefore \frac{s}{t} = r \frac{\theta}{t}.$$

But s/t = the *velocity*, v , of the rotating object, and θ/t is the angular velocity.

$$\therefore v = r\omega \quad (4)$$

Acceleration in a circle

When a stone is attached to a string and whirled round at constant speed in a circle, one can feel the force in the string needed to keep the stone moving. The presence of the force, called a *centripetal force*, implies that the stone has an acceleration. And since the force acts towards the centre of the circle, the direction of the acceleration, which is a vector quantity, is also towards the centre.

To obtain an expression for the acceleration towards the centre, consider an object moving with a constant speed v round a circle of radius r . Fig. 2.2 (i). At A, its velocity v_A is in the direction of the tangent AC; a short time δt later at B, its velocity v_B is in the direction of the tangent BD. Since their directions are different, the velocity v_B is different from the velocity v_A , although their magnitudes are both equal to v . Thus a velocity change or acceleration has occurred from A to B.

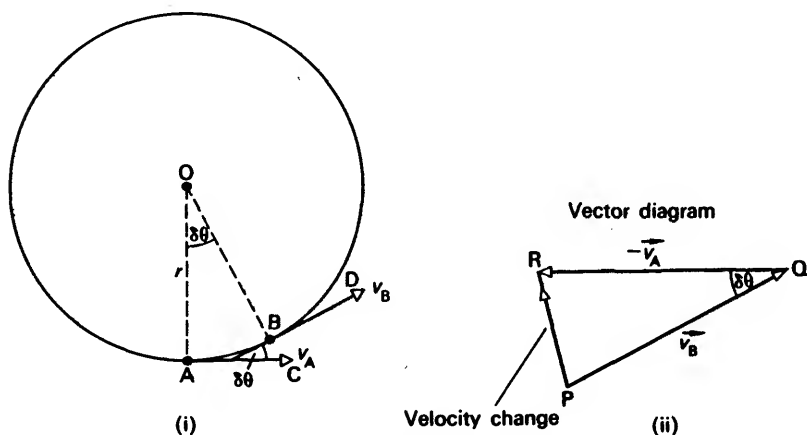


FIG. 2.2 Acceleration in circle

The velocity change from A to B $= \vec{v}_B - \vec{v}_A = \vec{v}_B + (-\vec{v}_A)$. The arrows denote vector quantities. In Fig. 2.2 (ii), PQ is drawn to represent v_B in magnitude (v) and direction (BD); QR is drawn to represent $(-\vec{v}_A)$ in magnitude (v) and direction (CA). Then, as shown on p. 11,

$$\text{velocity change} = \vec{v}_B + (-\vec{v}_A) = \vec{PR}.$$

When δt is small, the angle AOB or $\delta\theta$ is small. Thus angle PQR, equal to $\delta\theta$, is small. PR then points towards O, the centre of the circle. *The velocity change or acceleration is thus directed towards the centre.*

The magnitude of the acceleration, a , is given by

$$a = \frac{\text{velocity change}}{\text{time}} = \frac{PR}{\delta t} \\ = \frac{v \cdot \delta\theta}{\delta t}.$$

since $PR = v \cdot \delta\theta$. In the limit, when δt approaches zero, $\delta\theta/\delta t = d\theta/dt = \omega$, the angular velocity. But $v = r\omega$ (p. 36). Hence, since $a = v\omega$,

$$a = \frac{v^2}{r} \quad \text{or} \quad r\omega^2.$$

Thus an object moving in a circle of radius r with a constant speed v has a constant acceleration towards the centre equal to v^2/r or $r\omega^2$.

Centripetal forces

The force F required to keep an object of mass m moving in a circle of radius $r = ma = mv^2/r$. It is called a *centripetal force* and acts *towards the centre* of the circle. When a stone A is whirled in a horizontal circle of centre O by means of a string, the tension T provides the centripetal force. Fig. 2.3 (i). For a racing car moving round a circular track, the friction at the wheels provides the centripetal force. Planets such as P, moving in a circular orbit round the sun S, have a centripetal force due to gravitational attraction between S and P (p. 59). Fig. 2.3 (ii).

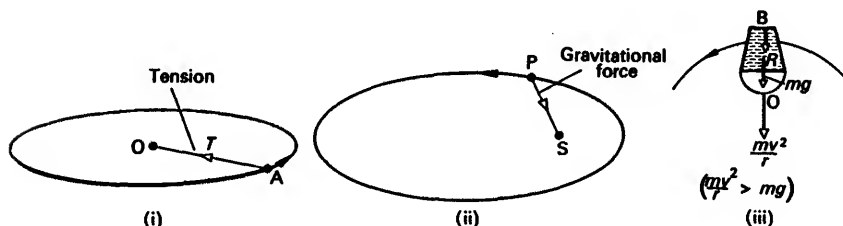


FIG. 2.3 Centripetal forces

If some water is placed in a bucket B attached to the end of a string, the bucket can be whirled in a vertical plane without any water falling out. When the bucket is vertically above the point of support O, the weight mg of the water is less than the required force mv^2/r towards the centre and so the water stays in. Fig. 2.3 (iii). The reaction R of the bucket base on the water provides the rest of the force. If the bucket is whirled slowly and $mg > mv^2/r$, part of the weight provides the force mv^2/r . The rest of the weight causes the water to accelerate downward and hence to leave the bucket.

Centrifuges

Centrifuges are used to separate particles in suspension from the less dense liquid in which they are contained. This mixture is poured into a tube in the centrifuge, which is then whirled at high speed in a horizontal circle.

The pressure gradient due to the surrounding liquid at a particular distance, r say, from the centre provides a centripetal force of $mr\omega^2$ for a small volume of liquid of mass m , where ω is the angular velocity.

If the volume of liquid is replaced by an equal volume of particles of smaller mass m' than the liquid, the centripetal force acting on the particles at the same place is then greater than that required by $(m-m')r\omega^2$. The net force urges the particles towards the centre in spiral paths, and here they collect. Thus when the centrifuge is stopped, and the container or tube assumes a vertical position, the suspension is found at the top of the tube and clear liquid at the bottom. For the same reason, cream is separated from the denser milk by spinning the mixture in a vessel. The cream spirals towards the centre and collects here.

Motion of Bicycle Rider Round Circular Track

When a person on a bicycle rides round a circular racing track, the frictional force F at the ground provides the inward force towards the centre or centripetal force. Fig. 2.4. This produces a moment about his centre of gravity G which is counterbalanced, when he leans inwards, by the moment of the normal reaction R . Thus provided no skidding occurs, $F \cdot h = R \cdot a = mg \cdot a$, since $R = mg$ for no vertical motion.

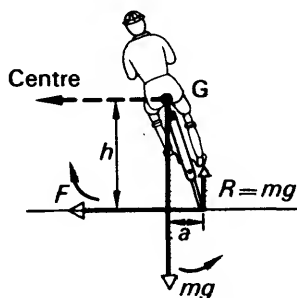


FIG. 2.4 Rider on circular track

$$\therefore \frac{a}{h} = \tan \theta = \frac{F}{mg},$$

where θ is the angle of inclination to the vertical. Now $F = mv^2/r$.

$$\therefore \tan \theta = \frac{v^2}{rg}.$$

When F is greater than the limiting friction, skidding occurs. In this case $F > \mu mg$, or $mg \tan \theta > \mu mg$. Thus $\tan \theta > \mu$ is the condition for skidding.

Motion of Car (or Train) Round Circular Track

Suppose a car (or train) is moving with a velocity v round a horizontal circular track of radius r , and let R_1 , R_2 be the respective normal re-

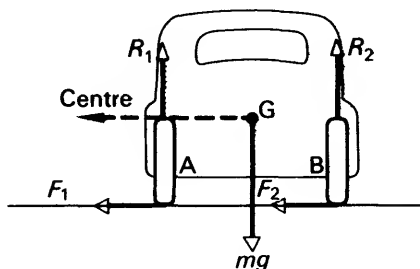


FIG. 2.5 Car on circular track

Thus for a given velocity v and radius r , the angle of inclination of the track for no side-slip must be $\tan^{-1}(v^2/rg)$. As the speed v increases, the angle θ increases, from (iii). A racing-track is made saucer-shaped because at higher speeds the cars can move towards a part of the track which is steeper and sufficient to prevent side-slip. The outer rail of a curved railway track is raised about the inner rail so that the force towards the centre is largely provided by the component of the reaction at the wheels. It is desirable to bank a road at corners for the same reason as a racing track is banked.

Thrust at Ground

Suppose now that the car (or train) is moving at such a speed that the frictional forces at A, B are F_1 , F_2 respectively, each acting towards the centre of the track. Resolving horizontally,

$$\therefore (R_1 + R_2) \sin \theta + (F_1 + F_2) \cos \theta = \frac{mv^2}{r} \quad (i)$$

Resolving vertically,

$$\therefore (R_1 + R_2) \cos \theta - (F_1 + F_2) \sin \theta = mg \quad (ii)$$

Solving, we find

$$F_1 + F_2 = m \left(\frac{v^2}{r} \cos \theta - g \sin \theta \right) \quad (iii)$$

If $\frac{v^2}{r} \cos \theta > g \sin \theta$, then $(F_1 + F_2)$ is positive; and in this case both the thrusts on the wheels at the ground are towards the centre of the track.

If $\frac{v^2}{r} \cos \theta < g \sin \theta$, then $(F_1 + F_2)$ is negative. In this case the forces F_1 and F_2 act outwards away from the centre of the track.

For stability, we have, by moments about G,

$$(F_1 + F_2)h + R_1a - R_2a = 0$$

$$\therefore (F_1 + F_2)\frac{h}{a} = R_2 - R_1.$$

From (iii),
$$\therefore \frac{mh}{a} \left(\frac{v^2}{r} \cos \theta - g \sin \theta \right) = R_2 - R_1 \quad (iv)$$

The reactions R_1 , R_2 can be calculated by finding $(R_1 + R_2)$ from equations (i), (ii), and combining the result with equation (iv). This is left as an exercise to the student.

Variation of g with latitude

The acceleration due to gravity, g , varies over the earth's surface. This is due to two main causes. Firstly, the earth is elliptical, with the polar radius, b , 6.357×10^6 metre and the equatorial radius, a , 6.378×10^6 metre, and hence g is greater at the poles than at the equator, where the body is further away from the centre of the earth. Secondly, the earth rotates about the polar axis, AB. Fig. 2.7. We shall consider this effect in more detail, and suppose the earth is a perfect sphere.

In general, an object of mass m suspended by a spring-balance at a

point on the earth would be acted on by an upward force $T = mg'$, where g' is the observed or apparent acceleration due to gravity. There would also be a downward attractive force mg towards the centre of the earth, where g is the acceleration in the absence of rotation.

(1) At the poles, A or B, there is no rotation. Hence $mg - T = 0$, or $mg = T = mg'$. Thus $g' = g$.

(2) At the equator, C or D, there is a resultant force $mr\omega^2$ towards the centre where r is the earth's radius. Since OD is the vertical, we have

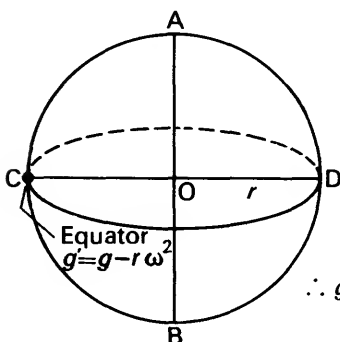


FIG. 2.7 Variation of g

$$mg - T = mr\omega^2.$$

$$\therefore T = mg - mr\omega^2 = mg'$$

$$\therefore g' = g - r\omega^2.$$

The radius r of the earth is about 6.37×10^6 m, and $\omega = [(2\pi)/(24 \times 3600)]$ radian per second.

$$\therefore g - g' = r\omega^2 = \frac{6.37 \times 10^6 \times (2\pi)^2}{(24 \times 3600)^2} = 0.034.$$

Latest figures give g , at the pole, 9.832 m s^{-2} , and g' , at the equator, 9.780 m s^{-2} , a difference of 0.052 m s^{-2} .

The earth's rotation accounts for 0.034 m s^{-2} .

EXAMPLE

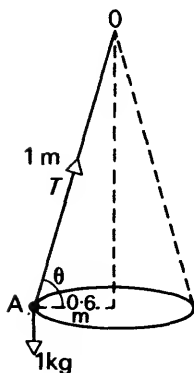


FIG. 2.8 Example

Explain the action of a centrifuge when used to hasten the deposition of a sediment from a liquid.

A pendulum bob of mass 1 kg is attached to a string 1 m long and made to revolve in a horizontal circle of radius 60 cm. Find the period of the motion and the tension of the string. (C.)

First part. See text, p. 38.

Second part. Suppose A is the bob, and OA is the string, Fig. 2.8. If T is the tension in newton, and θ is the angle of inclination of OA to the horizontal, then, for motion in the circle of radius $r = 60 \text{ cm} = 0.6 \text{ m}$,

$$T \cos \theta = \frac{mv^2}{r} = \frac{mv^2}{0.6} \quad \dots \quad (i)$$

Since the bob A does not move in a vertical direction, then

$$T \sin \theta = mg \quad \dots \quad (ii)$$

Now $\cos \theta = \frac{60}{100} = \frac{3}{5}$; hence $\sin \theta = \frac{4}{5}$.

From (ii),

$$\therefore T = \frac{mg}{\sin \theta} = \frac{1 \times 9.8}{4/5} = 12.25 \text{ newton.}$$

From (i)

$$v = \sqrt{\frac{0.6T \cos \theta}{m}}$$

$$= \sqrt{\frac{0.6 \times 12.25 \times 3}{1 \times 5}} = 2.1 \text{ m s}^{-1}$$

$$\therefore \text{angular velocity, } \omega = \frac{v}{r} = \frac{2.1}{0.6} = \frac{7}{2} \text{ rad s}^{-1}$$

$$\therefore \text{period, } T = \frac{2\pi}{\omega} = \frac{2\pi}{7/2} = \frac{4\pi}{7} \text{ second.}$$

$$\therefore T = 1.8 \text{ second.}$$

SIMPLE HARMONIC MOTION

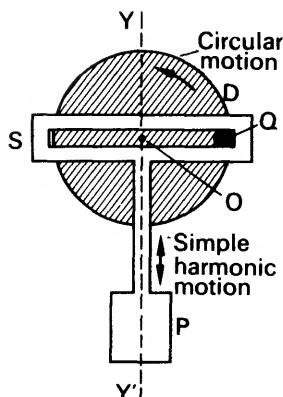


FIG. 2.9

Simple harmonic motion

When the bob of a pendulum moves to-and-fro through a small angle, the bob is said to be moving with *simple harmonic motion*. The prongs of a sounding tuning fork, and the layers of air near it, are moving with simple harmonic motion, and light waves can be considered due to simple harmonic variations.

Simple harmonic motion is closely associated with circular motion. An example is shown in Fig. 2.9. This illustrates an arrangement used to convert the circular motion of a disc D into the to-and-fro or simple harmonic motion of a piston P. The disc is driven about its axle O by a peg Q fixed near its rim. The vertical motion drives P up and down. Any horizontal component of the motion merely causes Q to move along the slot S. Thus the simple harmonic motion of P is the *projection* on the vertical line YY' of the circular motion of Q.

An everyday example of an opposite conversion of motion occurs in car engines. Here the to-and-fro or 'reciprocating' motion of the piston engine is changed to a regular circular motion by connecting rods and shafts so that the wheels are turned.

Formulae in Simple Harmonic Motion

Consider an object moving round a circle of radius r and centre Z with a uniform angular velocity ω , Fig. 2.10. If CZF is a fixed diameter, the foot of the perpendicular from the moving object to this diameter moves from Z to C, back to Z and across to F, and then returns to Z, while the object moves once round the circle from O in an anti-clockwise direction. The to-and-fro motion along CZF of the foot of the perpendicular is defined as *simple harmonic motion*.

Suppose the object moving round the circle is at A at some instant, where angle OZA = θ , and suppose the foot of the perpendicular from A to CZ is M. The acceleration of the object at A is $\omega^2 r$, and this

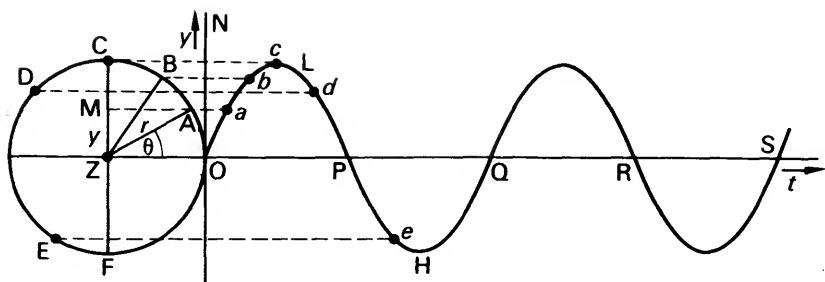


FIG. 2.10 Simple harmonic curve

acceleration is directed along the radius AZ (see p. 37). Hence the acceleration of M towards Z

$$= \omega^2 r \cos AZC = \omega^2 r \sin \theta.$$

But $r \sin \theta = MZ = y$ say.

$$\therefore \text{acceleration of } M \text{ towards } Z = \omega^2 y.$$

Now ω^2 is a constant.

$$\therefore \text{acceleration of } M \text{ towards } Z \propto \text{distance of } M \text{ from } Z.$$

If we wish to express mathematically that the acceleration is always directed towards Z , we must say

$$\text{acceleration towards } Z = -\omega^2 y \quad (1)$$

The minus indicates, of course, that the object begins to retard as it passes the centre, Z , of its motion. If the minus were omitted from equation (1) the latter would imply that the acceleration increases as y increases, and the object would then never return to its original position.

We can now form a definition of simple harmonic motion. It is the motion of a particle whose *acceleration is always (i) directed towards a fixed point, (ii) directly proportional to its distance from that point.*

Period, Amplitude. Sine Curve

The time taken for the foot of the perpendicular to move from C to F and back to C is known as the *period* (T) of the simple harmonic motion. In this time, the object moving round the circle goes exactly once round the circle from C ; and since ω is the angular velocity and 2π radians (360°) is the angle described, the period T is given by

$$T = \frac{2\pi}{\omega} \quad (1)$$

The distance ZC , or ZF , is the maximum distance from Z of the foot of the perpendicular, and is known as the *amplitude* of the motion. It is equal to r , the radius of the circle.

We have now to consider the variation with time, t , of the distance,

y , from Z of the foot of the perpendicular. The distance $y = ZM = r \sin \theta$. But $\theta = \omega t$, where ω is the angular velocity.

$$\therefore y = r \sin \omega t \quad (2)$$

The graph of y v. t is shown in Fig. 2.10, where ON represents the y -axis and OS the t -axis; since the angular velocity of the object moving round the circle is constant, θ is proportional to the time t . Thus as the foot of the perpendicular along CZF moves from Z to C and back to Z , the graph OLP is traced out; as the foot moves from Z to F and returns to Z , the graph PHQ is traced out. The graph is a *sine curve*. The complete set of values of y from O to Q is known as a cycle. The number of cycles per second is called the *frequency*. The unit '1 cycle per second' is called '1 hertz (Hz)'. The mains frequency in Great Britain is 50 Hz or 50 cycles per second.

Velocity during S.H.M.

Suppose the object moving round the circle is at A at some instant, Fig. 2.10. The velocity of the object is $r\omega$, where r is the radius of the circle, and it is directed along the tangent at A . Consequently the velocity parallel to the diameter FC at this instant $= r\omega \cos \theta$, by resolving.

$$\therefore \text{velocity, } v, \text{ of } M \text{ along } FC = r\omega \cos \theta.$$

But

$$y = r \sin \theta$$

$$\therefore \cos \theta = \sqrt{1 - \sin^2 \theta} = \sqrt{1 - y^2/r^2} = \frac{1}{r} \sqrt{r^2 - y^2}$$

$$\therefore v = \omega \sqrt{r^2 - y^2} \quad (1)$$

This is the expression for the velocity of an object moving with simple harmonic motion. The maximum velocity, v_m , corresponds to $y = 0$, and hence

$$v_m = \omega r. \quad (2)$$

Summarising our results:

(1) If the acceleration a of an object $= -\omega^2 y$, where y is the distance or displacement of the object from a fixed point, the motion is simple harmonic motion.

(2) The *period*, T , of the motion $= 2\pi/\omega$, where T is the time to make a complete to-and-fro movement or cycle. The *frequency*, f , $= 1/T$ and its unit is 'Hz'.

(3) The amplitude, r , of the motion is the maximum distance on either side of the centre of oscillation.

(4) The velocity at any instant, v , $= \omega \sqrt{r^2 - y^2}$; the maximum velocity $= \omega r$. Fig. 2.11 (i) shows a graph of the variation of v and acceleration a with displacement y , which are respectively an ellipse and a straight line.

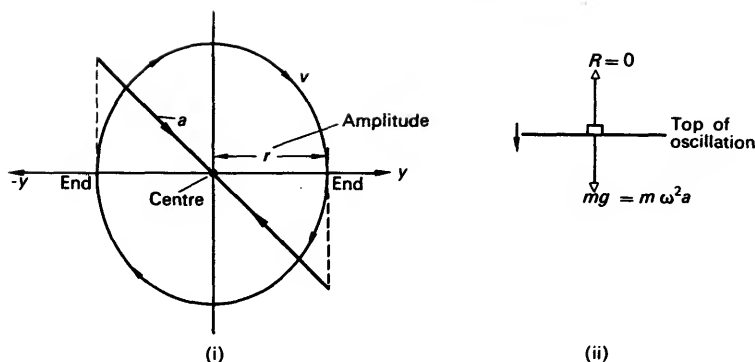


FIG. 2.11 Simple harmonic motion

S.H.M. and g

If a small coin is placed on a horizontal platform connected to a vibrator, and the amplitude is kept constant as the frequency is increased from zero, the coin will be heard 'chattering' at a particular frequency f_0 . At this stage the reaction of the table with the coin becomes zero at some part of every cycle, so that it loses contact periodically with the surface. Fig. 2.11 (ii).

The maximum acceleration in S.H.M. occurs at the end of the oscillation because the acceleration is directly proportional to the displacement. Thus maximum acceleration $= \omega^2 a$, where a is the amplitude and ω is $2\pi f_0$.

The coin will lose contact with the table when it is moving *down* with acceleration g (Fig. 2.11 (ii)). Suppose the amplitude is 8.0 cm. Then

$$(2\pi f_0)^2 a = g$$

$$\therefore 4\pi^2 f_0^2 \times 0.08 = 9.8$$

$$\therefore f_0 = \sqrt{\frac{9.8}{4\pi^2 \times 0.08}} = 1.8 \text{ Hz.}$$

Damping of S.H.M.

In practice, simple harmonic variations of a pendulum, for example, will die away as the energy is dissipated by viscous forces due to the air. The oscillation is then said to be *damped*. In the absence of any damping forces the oscillations are said to be *free*.

A simple experiment to investigate the effect of damping is illustrated in Fig. 2.12 (i). A suitable weight A is suspended from a helical spring S, a pointer P is attached to S, and a vertical scale R is set up behind P. The weight A is then set pulled down and released. The period, and the time taken for the oscillations to die away, are noted.

As shown in Fig. 2.12 (ii), A is now fully immersed in a damping medium, such as a light oil, water or glycerine. A is then set oscillating,

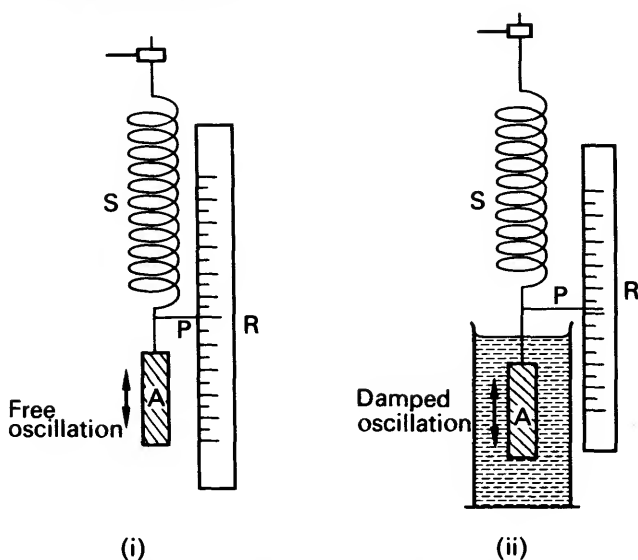


FIG. 2.12 Experiment on damped oscillations

and the time for oscillations to die away is noted. It is shorter than before and least for the case of glycerine. The decreasing amplitude in successive oscillations may also be noted from the upward limit of travel of P and the results plotted.

Fig. 2.13 (i), (ii) shows how damping produces an exponential fall in the amplitude with time.

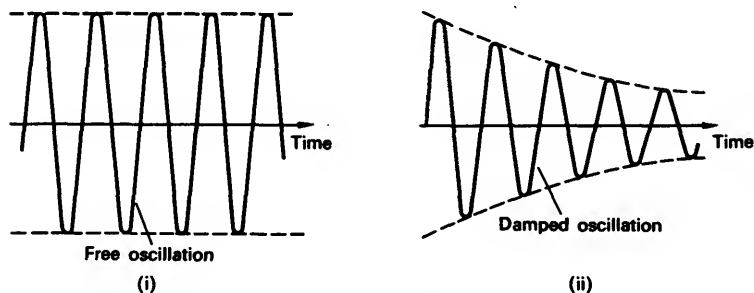


FIG. 2.13 Free and damped oscillations

The experiment works best for a period of about $\frac{1}{2}$ -second and a weight which is long and thin so that the damping is produced by non-turbulent fluid flow over the vertical sides. During the whole cycle, A must be totally immersed in the fluid.

EXAMPLE

A steel strip, clamped at one end, vibrates with a frequency of 20 Hz and an amplitude of 5 mm at the free end, where a small mass of 2 g is positioned.

Find (a) the velocity of the end when passing through the zero position, (b) the acceleration at maximum displacement, (c) the maximum kinetic and potential energy of the mass.

Suppose $y = r \sin \omega t$ represents the vibration of the strip where r is the amplitude.

(a) The velocity, $v = \omega \sqrt{r^2 - y^2}$ (p. 45). When the end of the strip passes through the zero position $y = 0$; and the maximum speed, v_m , is given by

$$v_m = \omega r.$$

Now $\omega = 2\pi f = 2\pi \times 20$, and $r = 0.005$ m.

$$\therefore v_m = 2\pi \times 20 \times 0.005 = 0.628 \text{ m s}^{-1}.$$

(b) The acceleration $= -\omega^2 y = -\omega^2 r$ at the maximum displacement.

$$\begin{aligned} \therefore \text{acceleration} &= (2\pi \times 20)^2 \times 0.005 \\ &= 79 \text{ m s}^{-2}. \end{aligned}$$

(c) $m = 2 \text{ g} = 2 \times 10^{-3} \text{ kg}$, $v_m = 0.628 \text{ m s}^{-1}$.

\therefore maximum K.E. $= \frac{1}{2} m v_m^2 = \frac{1}{2} \times (2 \times 10^{-3}) \times 0.628^2 = 3.9 \times 10^{-4} \text{ J (approx.)}$.

Maximum P.E. ($v = 0$) = Maximum K.E. $= 3.9 \times 10^{-4} \text{ J}$.

Simple Pendulum

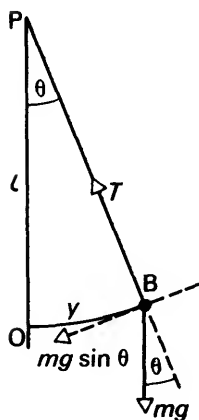


FIG. 2.14
Simple pendulum

We shall now study some cases of simple harmonic motion. Consider a *simple pendulum*, which consists of a small mass m attached to the end of a length l of wire, Fig. 2.14. If the other end of the wire is attached to a fixed point P and the mass is displaced slightly, it oscillates to-and-fro along the arc of a circle of centre P . We shall now show that the motion of the mass about its original position O is simple harmonic motion.

Suppose that the vibrating mass is at B at some instant, where $OB = y$ and angle $OPB = \theta$. At B , the force pulling the mass towards O is directed along the tangent at B , and is equal to $mg \sin \theta$. The tension, T , in the wire has no component in this direction, since PB is perpendicular to the tangent at B . Thus, since force = mass \times acceleration (p. 13),

$$-mg \sin \theta = ma,$$

where a is the acceleration along the arc OB ; the minus indicates that the force is towards O , while the displacement, y , is measured along the arc from O in the opposite direction. When θ is small, $\sin \theta = \theta$ in radians; also $\theta = y/l$. Hence,

$$-mg\theta = -mg \frac{y}{l} = ma$$

$$\therefore a = -\frac{g}{l}y = -\omega^2 y,$$

where $\omega^2 = g/l$. Since the acceleration is proportional to the distance y from a fixed point, the motion of the vibrating mass is simple harmonic motion (p. 50). Further, from p. 50, the period $T = 2\pi/\omega$.

$$\therefore T = \frac{2\pi}{\sqrt{g/l}} = 2\pi\sqrt{\frac{l}{g}} \quad (1)$$

At a given place on the earth, where g is constant, the formula shows that the period T depends only on the length, l , of the pendulum. Moreover, the period remains constant even when the amplitude of the vibrations diminish owing to the resistance of the air. This result was first obtained by Galileo, who noticed a swinging lantern one day, and timed the oscillations by his pulse. He found that the period remained constant although the swings gradually diminished in amplitude.

Determination of g by Simple Pendulum

The acceleration due to gravity, g , can be found by measuring the period, T , of a simple pendulum corresponding to a few different lengths, l , from 80 cm to 180 cm for example.

To perform the experiment accurately: (i) Fifty oscillations should be timed, (ii) a small angle of swing is essential, less than 10° , (iii) a small sphere should be tied to the end of a thread to act as the mass, and its radius added to the length of the thread to determine l .

A graph of l against T^2 is now plotted from the results, and a straight line AB, which should pass through the origin, is then drawn to lie evenly between the points, Fig. 2.15.

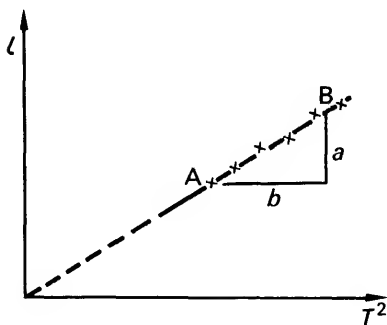


FIG. 2.15 Graph of l v. T^2

Now

$$T = 2\pi\sqrt{\frac{l}{g}}$$

$$\therefore T^2 = \frac{4\pi^2 l}{g}$$

$$\therefore g = 4\pi^2 \times \frac{l}{T^2} \quad (1)$$

The gradient a/b of the line AB is the magnitude of l/T^2 ; and by substituting in (1), g can then be calculated.

If the pendulum is suspended from the ceiling of a very tall room and the string and bob reaches nearly to the floor, then one may proceed to find g by (i) measuring the period T_1 , (ii) cutting off a measured length a of the string and determining the new period T_2 with the

shortened string. Then, if h is the height of the ceiling above the bob initially, $T_1 = 2\pi\sqrt{h/g}$ and $T_2 = 2\pi\sqrt{(h-a)/g}$. Thus

$$h = \frac{gT_1^2}{4\pi^2} \quad \text{and} \quad h-a = \frac{gT_2^2}{4\pi^2}.$$

$$\therefore a = \frac{g}{4\pi^2}(T_1^2 - T_2^2).$$

$$\therefore g = \frac{4\pi^2 a}{T_1^2 - T_2^2}.$$

Thus g can be calculated from a , T_1 and T_2 . Alternatively, the period T can be measured for several lengths a . Then, since $T = 2\pi\sqrt{(h-a)/g}$,

$$h-a = \frac{g}{4\pi^2}T^2.$$

A graph of a v. T^2 is thus a straight line whose gradient is $g/4\pi^2$. Hence g can be found. The intercept on the axis of a , when $T^2 = 0$, is h , the height of the ceiling above the bob initially.

The Spiral Spring or Elastic Thread

When a weight is suspended from the end of a spring or an elastic thread, experiment shows that the extension of the spring, i.e., the increase in length, is proportional to the weight, provided that the elastic limit of the spring is not exceeded (see p. 181). Generally, then, the tension (force), T , in a spring is proportional to the extension x produced, i.e., $T = kx$, where k is a constant of the spring.

Consider a spring or an elastic thread PA of length l suspended from a fixed point P, Fig. 2.16. When a mass m is placed on it, the spring stretches to O by a length e given by

$$mg = ke, \quad (i)$$

since the tension in the spring is then mg . If the mass is pulled down a little and then released, it vibrates up-and-down above and below O. Suppose at an instant that B is at a distance x below O. The tension T of the spring at B is then equal to $k(e+x)$, and hence the force towards O = $k(e+x) - mg$. Since force = mass \times acceleration,

$$\therefore -[k(a+x) - mg] = ma,$$

the minus indicates that the net force is upward at this instant, whereas the displacement x is measured from O in the opposite direction at the same instant. From this equation,

$$-ke - kx + mg = ma.$$

But, from (i),

$$mg = ke,$$

$$\therefore -kx = ma,$$

$$\therefore a = -\frac{k}{m}x = -\omega^2 x,$$

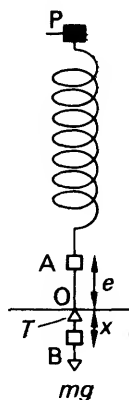


FIG. 2.16
Spiral spring

where $\omega^2 = k/m$. Thus the motion is simple harmonic about O, and the period T is given by

$$T = \frac{2\pi}{\omega} = 2\pi \sqrt{\frac{m}{k}} \quad (1)$$

Also, since $mg = ke$, it follows that $m/k = e/g$.

$$\therefore T = 2\pi \sqrt{\frac{e}{g}} \quad (2)$$

From (1), it follows that $T^2 = 4\pi^2 m/k$. Consequently a graph of T^2 v. m should be a straight line passing through the origin. In practice, when the load m is varied and the corresponding period T is measured, a straight line graph is obtained when T^2 is plotted against m , thus verifying indirectly that the motion of the load was simple harmonic. The graph does not pass through the origin, however, owing to the mass and the movement of the various parts of the spring. This has not been taken into account in the foregoing theory and we shall now show how g may be found in this case.

Determination of g by Spiral Spring

The mass s of a vibrating spring is taken into account, in addition to the mass m suspended at the end, theory beyond the scope of this book then shows that the period of vibration, T , is given by

$$T = 2\pi \sqrt{\frac{m + \lambda s}{k}} \quad (i)$$

where λ is approximately $\frac{1}{3}$ and k is the elastic constant of the spring. Squaring (i) and re-arranging,

$$\therefore \frac{k}{4\pi^2} T^2 = m + \lambda s \quad (ii)$$

Thus, since λ , k , s are constants, a graph of T^2 v. m should be a straight line when m is varied and T observed. A straight line graph verifies indirectly that the motion of the mass at the end of the spring is simple harmonic. Further, the magnitude of $k/4\pi^2$ can be found from the slope of the line, and hence k can be calculated.

If a mass M is placed on the end of the spring, producing a steady extension e less than the elastic limit, then $Mg = ke$.

$$\therefore g = \frac{e}{M} \times k \quad (iii)$$

By attaching different masses to the spring, and measuring the corresponding extension, the magnitude of e/M can be found by plotting e v. M and measuring the slope of the line. This is called the 'static' experiment on the spring. From the magnitude of k obtained in the 'dynamic' experiment when the period was determined for different loads, the value of g can be found by substituting the magnitudes of e/M and k in (iii).

Oscillations of a Liquid in a U-Tube

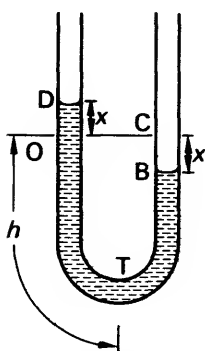


FIG. 2.17
S.H.M. of liquid

If the liquid on one side of a U-tube T is depressed by blowing gently down that side, the levels of the liquid will oscillate for a short time about their respective initial positions O, C, before finally coming to rest, Fig. 2.17.

The period of oscillation can be found by supposing that the level of the liquid on the left side of T is at D at some instant, at a height x above its original (undisturbed) position O. The level B of the liquid on the other side is then at a depth x below its original position C, and hence the excess pressure on the whole liquid, as shown on p. 110,

$$= \text{excess height} \times \text{density of liquid} \times g \\ = 2x\rho g.$$

Now pressure = force per unit area.

$$\therefore \text{force on liquid} = \text{pressure} \times \text{area of cross-section of the tube} \\ = 2x\rho g \times A,$$

where A is the cross-sectional area of the tube.

This force causes the liquid to accelerate. The mass of liquid in the U-tube = volume \times density = $2hA\rho$, where $2h$ is the total length of the liquid in T. Now the acceleration, a , towards O or C is given by $\text{force} = \text{mass} \times a$.

$$\therefore -2x\rho gA = 2hA\rho a.$$

The minus indicates that the force towards O is opposite to the displacement measured from O at that instant.

$$\therefore a = -\frac{g}{h}x = -\omega^2 x,$$

where $\omega^2 = \frac{g}{h}$. The motion of the liquid about O (or C) is thus simple harmonic, and the period T is given by

$$T = \frac{2\pi}{\omega} = 2\pi \sqrt{\frac{h}{g}}.$$

P.E. and K.E. exchanges in oscillating systems

We can now make a general point about *oscillations* and *oscillating systems*. As an illustration, suppose that one end of a spring S of negligible mass is attached to a smooth object A, and that S and A are laid on a horizontal smooth table. If the free end of S is attached to the table and A is pulled slightly to extend the spring and then released, the system vibrates with simple harmonic motion. This is the case discussed on p. 50, without taking gravity into account. The centre of oscillation O is the position of the end of the spring corresponding

to its natural length, that is, when the spring is neither extended or compressed. If the spring extension obeys the law $\text{force} = kx$, where k is a constant, and m is the mass of A, then, as on p. 51, it can easily be shown that the period T of oscillation is given by:

$$T = \frac{2\pi}{\omega} = 2\pi \sqrt{\frac{m}{k}}.$$

The energy of the stretched spring is *potential energy*, P.E.—its molecules are continually displaced or compressed relative to their normal distance apart. The P.E. for an extension $x = \int F \cdot dx = \int kx \cdot dx = \frac{1}{2}kx^2$.

The energy of the mass is *kinetic energy*, K.E., or $\frac{1}{2}mv^2$, where v is the velocity. Now from $x = a \sin \omega t$, $v = dx/dt = \omega a \cos \omega t$.

$$\begin{aligned} \therefore \text{total energy of spring plus mass} &= \frac{1}{2}kx^2 + \frac{1}{2}mv^2 \\ &= \frac{1}{2}ka^2 \sin^2 \omega t + \frac{1}{2}m\omega^2 a^2 \cos^2 \omega t. \end{aligned}$$

But $\omega^2 = k/m$, or $k = m\omega^2$.

$$\therefore \text{total energy} = \frac{1}{2}m\omega^2 a^2 (\sin^2 \omega t + \cos^2 \omega t) = \frac{1}{2}m\omega^2 a^2 = \text{constant}.$$

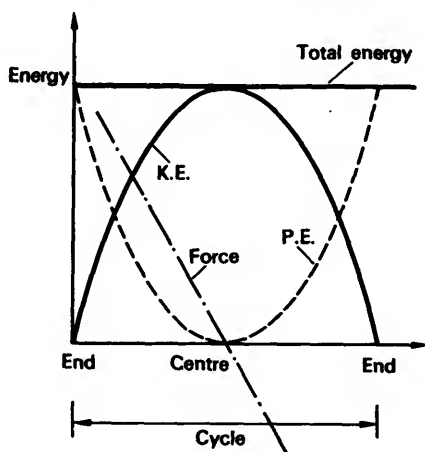


FIG. 2.18 Energy of S.H.M.

Thus the total energy of the vibrating mass and spring is constant. When the K.E. of the mass is a maximum (energy = $\frac{1}{2}m\omega^2 a^2$ and mass passing through the centre of oscillation), the P.E. of the spring is then zero ($x = 0$). Conversely, when the P.E. of the spring is a maximum (energy = $\frac{1}{2}ka^2 = \frac{1}{2}m\omega^2 a^2$ and mass at end of the oscillation), the K.E. of the mass is zero ($v = 0$). Fig. 2.18 shows the variation of P.E. and K.E. with displacement x ; the force F extending the spring, also shown, is directly proportional to the displacement from the centre of oscillation.

The constant interchange of energy between potential and kinetic energies is essential for producing and maintaining oscillations, whatever their nature. In the case of the oscillating bob of a simple pendulum,

for example, the bob loses kinetic energy after passing through the middle of the swing, and then stores the energy as potential energy as it rises to the top of the swing. The reverse occurs as it swings back. In the case of oscillating layers of air when a sound wave passes, kinetic energy of the moving air molecules is converted to potential energy when the air is compressed. In the case of electrical oscillations, a coil L and a capacitor C in the circuit constantly exchange energy; this is stored alternately in the magnetic field of L and the electric field of C .

EXAMPLES

1. Define *simple harmonic motion* and state the relation between displacement from its mean position and the restoring force when a body executes simple harmonic motion.

A body is supported by a spiral spring and causes a stretch of 1.5 cm in the spring. If the mass is now set in vertical oscillation of small amplitude, what is the periodic time of the oscillation? (L .)

First part. Simple harmonic motion is the motion of an object whose acceleration is proportional to its distance from a fixed point and is always directed towards that point. The relation is: Restoring force = $-k \times$ distance from fixed point, where k is a constant.

Second part. Let m be the mass of the body in kg. Then, since 1.5 cm = 0.015 m

$$mg = k \times 0.015 \quad \dots \quad (i)$$

where k is a constant of the spring in N m^{-1} . Suppose the vibrating body is x m below its original position at some instant and is moving downwards. Then since the extension is $(x + 0.015)$ m, the net downward force

$$\begin{aligned} &= mg - k(x + 0.015) \\ &= mg - k \times 0.015 - kx = -kx \end{aligned}$$

from (i). Now mass \times acceleration = force.

$$\therefore m \times \text{acceleration} = -kx$$

$$\therefore \text{acceleration} = -\frac{k}{m}x.$$

But, from (i),
$$\frac{k}{m} = \frac{g}{0.015}$$

$$\therefore \text{acceleration} = \frac{g}{0.015}x = -\omega^2 x,$$

where $\omega^2 = g/0.015$.

$$\begin{aligned} \therefore \text{period } T &= \frac{2\pi}{\omega} = 2\pi \sqrt{\frac{0.015}{g}} = 2\pi \sqrt{\frac{0.015}{9.8}} \\ &= 0.25 \text{ second.} \end{aligned}$$

2. A small bob of mass 20 g oscillates as a simple pendulum, with amplitude 5 cm and period 2 seconds. Find the velocity of the bob and the tension in the supporting thread, when the velocity of the bob is a maximum.

First part. See text.

Second part. The velocity, v , of the bob is a maximum when it passes through its original position. With the usual notation (see p. 45), the maximum velocity v_m is given by

$$v_m = \omega r,$$

where r is the amplitude of 0.05 m. Since $T = 2\pi/\omega$,

$$\therefore \omega = \frac{2\pi}{T} = \frac{2\pi}{2} = \pi \quad \dots \dots \dots (i)$$

$$\therefore v_m = \omega a = \pi \times 0.05 = 0.16 \text{ m s}^{-1}.$$

Suppose P is the tension in the thread. The net force towards the centre of the circle along which the bob moves is then given by $(P - mg)$. The acceleration towards the centre of the circle, which is the point of suspension, is v_m^2/l , where l is the length of the pendulum.

$$\therefore P - mg = \frac{mv_m^2}{l}$$

$$\therefore P = mg + \frac{mv_m^2}{l}$$

Now

$$T = 2\pi \sqrt{\frac{l}{g}}$$

$$\therefore l = \frac{gT^2}{4\pi^2} = \frac{g \times 4}{4\pi^2} = \frac{g}{\pi^2}$$

Since $m = 0.02 \text{ kg}$, $g = 9.8 \text{ m s}^{-2}$, it follows from above that

$$\begin{aligned} P &= 0.02 \times 9.8 + \frac{0.02 \times (0.05\pi)^2 \times \pi^2}{9.8} \\ &= 19.65 \times 10^{-2} \text{ newton} \end{aligned}$$

Waves. Wave equation

Waves and their properties can be demonstrated by producing them on the surface of water, as in a ripple tank. As the wave travels outwards from the centre of disturbance, it reaches more distant particles of water at a later time. Thus the particles of water vibrate out of *phase* with each other while the wave travels. It should be noted that the vibrating particles are the origin of the wave. Their mean position remains the same as the wave travels, but like the simple harmonic oscillators previously discussed, they store and release *energy* which is handed on from one part of the medium to another. The wave shows the energy travelling through the medium.

If the displacement y of a vibrating particle P is represented by $y = a \sin \omega t$, the displacement of a neighbouring particle Q can be represented by $y = a \sin (\omega t + \phi)$. ϕ is called the *phase angle* between the two vibrations. If $\phi = \pi/2$ or 90° , the vibration of Q is $y = a \sin (\omega t + \pi/2)$. In this case, $y = 0$ when $t = 0$ for P, but $y = a \sin \pi/2 = a$ when $t = 0$ for Q. Comparing the two simple harmonic variations, it can be seen that Q *leads* on P by a quarter of a period.

If the wave is 'frozen' at different times, the displacements of the various particles will vary according to their position or distance x

from some chosen origin such as the centre of disturbance. Now the *wavelength*, λ , of a wave is the distance between successive crests or troughs. At these points the phase difference is 2π . Consequently the phase angle for a distance x is $(x/\lambda) \times 2\pi$ or $2\pi x/\lambda$. The *wave equation*, which takes x into account as well as the time t , can thus be written as:

$$y = a \sin \left(2\pi \frac{t}{T} - 2\pi \frac{x}{\lambda} \right) = a \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda} \right) \quad (1)$$

Other forms of the wave equation may be used. The velocity v of a wave is the distance travelled by the disturbance in 1 second. If the frequency of the oscillations is f , then f waves travel outwards in 1 second. Each wave occupies a length λ . Hence $v = f\lambda$. Further, the period T is the time for 1 oscillation. Thus $f = 1/T$ and hence $v = f\lambda = \lambda/T$. Substituting for T in (1), the wave equation may also be written as:

$$y = a \sin \frac{2\pi}{\lambda} (vt - x) \quad (2)$$

The wave equation in (1) or (2) is a *progressive wave*. The energy of the wave travels outwards through the medium as time goes on.

Longitudinal and transverse waves

Waves can be classified according to the direction of their vibrations. A *longitudinal wave* is one produced by *vibrations parallel* to the direction of travel of the wave. An example is a sound wave. The layers of air are always vibrating in a direction parallel to the direction of travel of the wave. A longitudinal wave can be seen travelling in a 'Slinky' coil when one end is fixed and the other is pulled to-and-fro in the direction of the coil.

A *transverse wave* is one produced by *vibrations perpendicular* to the direction of travel of the wave. Light waves are transverse waves. The wave along a bowed string of a violin is a transverse wave.

Velocity of waves

There are various types of waves. A longitudinal wave such as a sound wave is a *mechanical wave*. The speed v with which the energy travels depends on the restoring stress after particles in the medium are strained from their original position. Thus v depends on the *modulus of elasticity* of the medium. It also depends on the inertia of the particles, of which the mass per unit volume or density ρ is a measure.

By dimensions, as well as rigorously, it can be shown that

$$v = \sqrt{\frac{\text{modulus of elasticity}}{\rho}}$$

For a solid, the modulus is Young's modulus, E . Thus $v = \sqrt{E/\rho}$. For a liquid or gas, the modulus is the bulk modulus, k . Hence $v = \sqrt{k/\rho}$. In air, $k = \gamma p$, where γ is the ratio of the principal specific heats of air and p is the atmospheric pressure. Thus $v = \sqrt{\gamma p/\rho}$ (p. 163).

When a taut string is plucked or bowed, the velocity of the transverse wave along it is given by $v = \sqrt{T/m}$, where T is the tension and m is the

mass per unit length of the string. In this case T provides the restoring force acting on the displaced particles of string and m is a measure of their inertia.

Electromagnetic waves, which are due to electric and magnetic vibrations, form an important group of waves in nature. Radio waves, infra-red, visible and ultra-violet light, X-rays, and γ -rays are all electromagnetic waves, ranging from long wavelengths such as 1000 metres (radio waves) to short wavelengths such as 10^{-8} m (γ -waves). Unlike the mechanical waves, no material medium is needed to carry the waves. The speed of all electromagnetic waves in a vacuum is the same, about 3×10^8 metre per second. The speed varies with wavelength in material media and this explains why dispersion (separation of colours) of white light is produced by glass.

Stationary waves

The equation $y = a \sin 2\pi(t/T - x/\lambda)$ represents a progressive wave travelling in the x -direction. A wave of the same amplitude and frequency travelling in the *opposite direction* is represented by the same form of equation but with $-x$ in place of x , that is, by $y = a \sin 2\pi(t/T + x/\lambda)$.

The *principle of superposition* states that the combined effect or resultant of two waves in a medium can be obtained by adding the displacements at each point due to the respective waves. Thus if the displacement due to one wave is represented by y_1 , and that due to the other wave by y_2 , the resultant displacement y is given by

$$\begin{aligned} y &= y_1 + y_2 = a \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda} \right) + a \sin 2\pi \left(\frac{t}{T} + \frac{x}{\lambda} \right) \\ &= 2a \sin 2\pi \frac{t}{T} \cdot \cos 2\pi \frac{x}{\lambda} = A \sin 2\pi \frac{t}{T}, \end{aligned}$$

where $A = 2a \cos 2\pi x/\lambda$.

A represents the amplitude at different points in the medium. When $x = 0$, $y = A$; when $x = \lambda/4$, $A = 0$; when $x = \lambda/2$, $y = -A$; when $x = 3\lambda/4$, $y = 0$. Thus at some points called *antinodes*, A , the amplitude of vibration is a maximum. At points half-way between the antinodes called *nodes*, N , the amplitude is zero, that is, there is no vibration here. Fig. 2.19 (i). This type of wave, which stays in one place in a medium, is called a *stationary* or *standing wave*. Stationary waves may be produced which are either longitudinal or transverse.

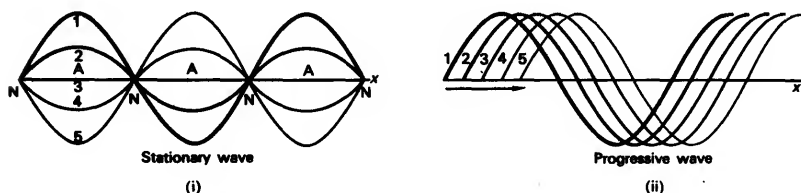


FIG. 2.19 Stationary and progressive waves

Unlike the progressive wave, where the energy travels outwards through the medium, Fig. 2.19 (ii), the energy of the stationary wave remains stored in one part of the medium. Stationary waves are produced in musical instruments when they are played. Stationary radio waves are also produced in receiving aerials. Stationary waves, due to electron motion, are believed to be present around the nucleus of atoms.

Interference. Diffraction

A stationary wave is a special case of *interference* between two waves. Another example occurs when two tuning forks of nearly equal frequency are sounded together. A periodic variation of loud sounds called 'beats' is then heard. They are due to the periodic variation of the amplitude of the resultant wave. If two very close coherent sources of light are obtained, interference between the two waves may produce bright and dark bands.

Diffraction is the name given to the interference between waves coming from coherent sources on the same undivided wavefront. The effect is pronounced when a wave is incident on a narrow opening whose width is of comparable order to the wavelength. The wave now spreads out or is 'diffracted' after passing through the slit. If the width of the slit, however, is large compared with the wavelength, the wave passes straight through the opening without any noticeable diffraction. This is why visible light, which has wavelengths of the order of 6×10^{-7} m, passes straight through wide openings and produces sharp shadows; whereas sound, which has wavelengths over a million times longer and of the order of say 0.5 m, can be heard round corners.

Further details of wave phenomena are discussed in the Sound and Optics sections of the book.

GRAVITATION

Kepler's Laws

The motion of the planets in the heavens had excited the interest of the earliest scientists, and Babylonian and Greek astronomers were able to predict their movements fairly accurately. It was considered for some time that the earth was the centre of the universe, but about 1542 COPERNICUS suggested that the planets revolved round the sun as centre. A great advance was made by KEPLER about 1609. He had studied for many years the records of observations on the planets made by TYCHO BRAHE, and he enunciated three laws known by his name. These state:

- (1) The planets describe ellipses about the sun as one focus.
- (2) The line joining the sun and the planet sweeps out equal areas in equal times.

(3) The squares of the periods of revolution of the planets are proportional to the cubes of their mean distances from the sun.

The third law was announced by Kepler in 1619.

Newton's Law of Gravitation

About 1666, at the early age of 24, NEWTON discovered a universal law known as the *law of gravitation*.

He was led to this discovery by considering the motion of a planet moving in a circle round the sun S as centre. Fig. 2.20 (i). The force acting on the planet of mass m is $m r \omega^2$, where r is the radius of the circle and ω is the angular velocity of the motion (p. 38). Since $\omega = 2\pi/T$, where T is the period of the motion,

$$\text{force on planet} = m r \left(\frac{2\pi}{T} \right)^2 = \frac{4\pi^2 m r}{T^2}.$$

This is equal to the force of attraction of the sun on the planet. Assuming an inverse-square law, then, if k is a constant,

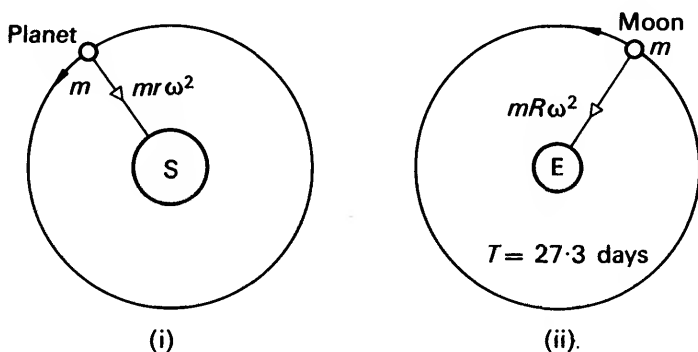


FIG. 2.20 Satellites

$$\text{force on planet} = \frac{k m}{r^2}.$$

$$\therefore \frac{k m}{r^2} = \frac{4\pi^2 m r}{T^2}$$

$$\therefore T^2 = \frac{4\pi^2}{k} r^3$$

$$\therefore T^2 \propto r^3,$$

since k, π are constants.

Now Kepler had announced that the squares of the periods of revolution of the planets are proportional to the cubes of their mean distances from the sun (see above). Newton thus suspected that *the force between the sun and the planet was inversely proportional to the square of the distance between them*. The great scientist now proceeded to test the inverse-square law by applying it to the case of the moon's motion

round the earth. Fig. 2.20 (ii). The moon has a period of revolution, T , about the earth of approximately 27.3 days, and the force on it $= mR\omega^2$, where R is the radius of the moon's orbit and m is its mass.

$$\therefore \text{force} = mR\left(\frac{2\pi}{T}\right)^2 = \frac{4\pi^2 mR}{T^2}.$$

If the planet were at the earth's surface, the force of attraction on it due to the earth would be mg , where g is the acceleration due to gravity. Fig. 2.20 (ii). Assuming that the force of attraction varies as the inverse square of the distance between the earth and the moon,

$$\therefore \frac{4\pi^2 mR}{T^2} : mg = \frac{1}{R^2} : \frac{1}{r^2},$$

where r is the radius of the earth.

$$\begin{aligned}\therefore \frac{4\pi^2 R}{T^2 g} &= \frac{r^2}{R^2}, \\ \therefore g &= \frac{4\pi^2 R^3}{r^2 T^2} \quad \dots \dots \dots (1)\end{aligned}$$

Newton substituted the then known values of R , r , and T , but was disappointed to find that the answer for g was not near to the observed value, 9.8 m s^{-2} . Some years later, he heard of a new estimate of the radius of the moon's orbit, and on substituting its value he found that the result for g was close to 9.8 m s^{-2} . Newton saw that a universal law could be formulated for the attraction between any two particles of matter. He suggested that: *The force of attraction between two given masses is inversely proportional to the square of their distance apart.*

Gravitational Constant, G , and its Determination

From Newton's law, it follows that the force of attraction, F , between two masses m , M at a distance r apart is given by $F \propto \frac{mM}{r^2}$:

$$\therefore F = G \frac{mM}{r^2}, \quad \dots \dots \dots (2)$$

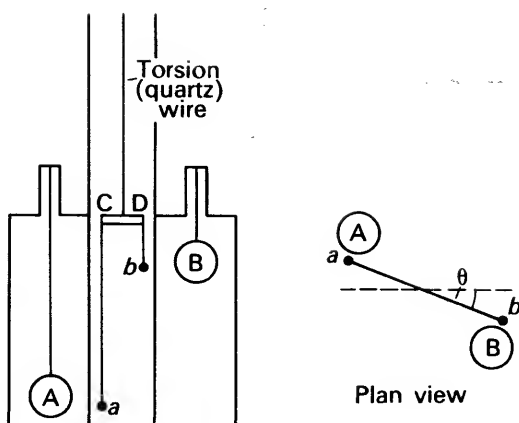
where G is a universal constant known as the *gravitational constant*. This expression for F is *Newton's law of gravitation*.

From (2), it follows that G can be expressed in ' $\text{N m}^2 \text{ kg}^{-2}$ '. The dimensions of G are given by

$$[G] = \frac{\text{MLT}^{-2} \times \text{L}^2}{\text{M}^2} = \text{M}^{-1} \text{L}^3 \text{T}^{-2}.$$

Thus the unit of G may also be expressed as $\text{m}^3 \text{ kg}^{-1} \text{ s}^{-2}$.

A celebrated experiment to measure G was carried out by C. V. BOYS in 1895, using a method similar to one of the earliest determinations of G by CAVENDISH in 1798. Two identical balls, a , b , of gold,

FIG. 2.21 Experiment on G

5 mm in diameter, were suspended by a long and a short fine quartz fibre respectively from the ends, C, D, of a highly-polished bar CD, Fig. 2.21. Two large identical lead spheres, A, B, 115 mm in diameter, were brought into position near a , b respectively. As a result of the attraction between the masses, two equal but opposite forces acted on CD. The bar was thus deflected, and the angle of deflection, θ , was measured by a lamp and scale method by light reflected from CD. The high sensitivity of the quartz fibres enabled the small deflection to be measured accurately, and the small size of the apparatus allowed it to be screened considerably from air convection currents.

Calculation for G

Suppose d is the distance between a , A, or b , B, when the deflection is θ . Then if m , M are the respective masses of a , A,

$$\text{torque of couple on CD} = G \frac{mM}{d^2} \times CD.$$

But torque of couple = $c\theta$,

where c is the torque in the torsion wire per unit radian of twist (p. 192).

$$\therefore G \frac{mM}{d^2} \times CD = c\theta.$$

$$\therefore G = \frac{c\theta d^2}{mM \times CD} \quad (1)$$

The constant c was determined by allowing CD to oscillate through a small angle and then observing its period of oscillation, T , which was of the order of 3 minutes. If I is the known moment of inertia of the system about the torsion wire, then (see p. 75),

$$T = 2\pi \sqrt{\frac{I}{c}}.$$

The constant c can now be calculated, and by substitution in (i), G can be determined. Accurate experiments showed that $G = 6.66 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$ and Heyl, in 1942, found G to be $6.67 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$.

Mass and Density of Earth

At the earth's surface the force of attraction on a mass m is mg , where g is the acceleration due to gravity. Now it can be shown that it is legitimate in calculations to assume that the mass, M , of the earth is concentrated at its centre, if it is a sphere. Assuming that the earth is spherical and of radius r , it then follows that the force of attraction of the earth on the mass m is GmM/r^2 .

$$\therefore G \frac{mM}{r^2} = mg.$$

$$\therefore g = \frac{GM}{r^2}.$$

$$\therefore M = \frac{gr^2}{G}.$$

Now, $g = 9.8 \text{ m s}^{-2}$, $r = 6.4 \times 10^6 \text{ m}$, $G = 6.7 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$.

$$\therefore M = \frac{9.8 \times (6.4 \times 10^6)^2}{6.7 \times 10^{-11}} = 6.0 \times 10^{24} \text{ kg}.$$

The volume of a sphere is $4\pi r^3/3$, where r is its radius. Thus the density, ρ , of the earth is approximately given by

$$\rho = \frac{M}{V} = \frac{gr^2}{4\pi r^3 G/3} = \frac{3g}{4\pi r G}.$$

By substituting known values of g , G , and r , the mean density of the earth is found to be about 5500 kg m^{-3} . The density may approach a value of 10000 kg m^{-3} towards the interior.

It is now believed that gravitational force travels with the speed of light. Thus if the gravitational force between the sun and earth were suddenly to disappear by the vanishing of the sun, it would take about 7 minutes for the effect to be experienced on the earth. The earth would then fly off along a tangent to its original curved path.

Gravitational and inertial mass

The mass m of an object appearing in the expression $F = ma$, force = mass \times acceleration, is the *inertial mass*, as stated on p. 13. It is a measure of the reluctance of the object to move when forces act on it. It appears in $F = ma$ from Newton's second law of motion.

The 'mass' of the same object concerned in Newton's theory of gravitational attraction can be distinguished from the inertial mass. This is called the *gravitational mass*. If it is given the symbol m_g , then $F_g = GMm_g/r^2$, where F_g is the gravitational force, M is the mass of the earth and r its radius. Now $GM/r^2 = g$, the acceleration due to gravity (see above). Thus $F_g = m_g g = W$, the weight of the object.

In the simple pendulum theory on p. 48, we can derive the period T using $W = \text{weight} = m_g g$ in place of the symbols adopted there.

Thus
$$-m_g g \frac{y}{l} = ma,$$

or
$$a = -\frac{m_g g}{ml} \cdot y = -\omega^2 y.$$

$$\therefore T = \frac{2\pi}{\omega} = 2\pi \sqrt{\frac{ml}{m_g g}}.$$

Experiments show that to a high degree of accuracy, $T = 2\pi\sqrt{l/g}$ no matter what mass is used, that is, the period depends only on l and g . Thus $m = m_g$, or the gravitational mass is equal to the inertial mass to the best of our present knowledge.

Mass of Sun

The mass M_s of the sun can be found from the period of a satellite and its distance from the sun. Consider the case of the earth. Its period T is about 365 days or $365 \times 24 \times 3600$ seconds. Its distance r_s from the centre of the sun is about 1.5×10^{11} m. If the mass of the earth is m , then, for circular motion round the sun,

$$\frac{GM_s m}{r_s^2} = m r_s \omega^2 = \frac{m r_s 4\pi^2}{T^2},$$

$$\therefore M_s = \frac{4\pi^2 r_s^3}{G T^2} = \frac{4\pi^2 \times (1.5 \times 10^{11})^3}{6.7 \times 10^{-11} \times (365 \times 24 \times 3600)^2} = 2 \times 10^{30} \text{ kg}.$$

Orbits round the earth

Satellites can be launched from the earth's surface to circle the earth. They are kept in their orbit by the gravitational attraction of the earth.

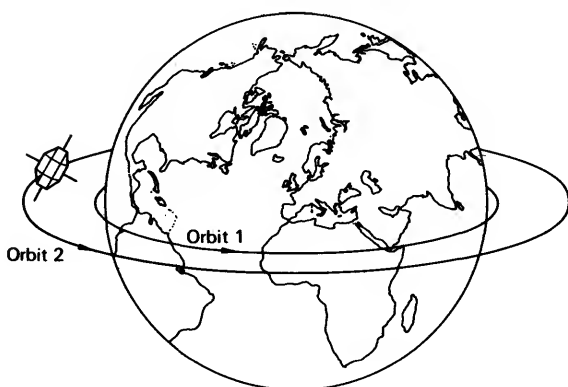


FIG. 2.22 Orbits round earth

Consider a satellite of mass m which just circles the earth of mass M

close to its surface in an orbit 1. Fig. 2.22 (i). Then, if r is the radius of the earth,

$$\frac{mv^2}{r} = G \frac{Mm}{r^2} = mg,$$

where g is the acceleration due to gravity at the earth's surface and v is the velocity of m in its orbit. Thus $v^2 = rg$, and hence, using $r = 6.4 \times 10^6$ m and $g = 9.8$ m s⁻²,

$$\begin{aligned} v &= \sqrt{rg} = \sqrt{6.4 \times 10^6 \times 9.8} = 8 \times 10^3 \text{ m s}^{-1} \text{ (approx),} \\ &= 8 \text{ km s}^{-1}. \end{aligned}$$

The velocity v in the orbit is thus about 8 km s⁻¹. In practice, the satellite is carried by a rocket to the height of the orbit and then given an impulse, by firing jets, to deflect it in a direction parallel to the tangent of the orbit (see p. 66). Its velocity is boosted to 8 km s⁻¹ so that it stays in the orbit. The period in orbit

$$\begin{aligned} &= \frac{\text{circumference of earth}}{v} = \frac{2\pi \times 6.4 \times 10^6 \text{ m}}{8 \times 10^3 \text{ m s}^{-1}} \\ &= 5000 \text{ seconds (approx)} = 83 \text{ min.} \end{aligned}$$

Parking Orbits

Consider now a satellite of mass m circling the earth in the plane of the equator in an orbit 2 concentric with the earth. Fig. 2.22 (ii). Suppose the direction of rotation as the same as the earth and the orbit is at a distance R from the centre of the earth. Then if v is the velocity in orbit,

$$\frac{mv^2}{R} = \frac{GMm}{R^2}.$$

But $GM = gr^2$, where r is the radius of the earth.

$$\begin{aligned} \therefore \frac{mv^2}{R} &= \frac{mgr^2}{R^2} \\ \therefore v^2 &= \frac{gr^2}{R}. \end{aligned}$$

If T is the period of the satellite in its orbit, then $v = 2\pi R/T$.

$$\begin{aligned} \therefore \frac{4\pi^2 R^2}{T^2} &= \frac{gr^2}{R} \\ \therefore T^2 &= \frac{4\pi^2 R^3}{gr^2}. \end{aligned} \quad (i)$$

If the period of the satellite in its orbit is exactly equal to the period of the earth as it turns about its axis, which is 24 hours, *the satellite will stay over the same place on the earth while the earth rotates.* This

is sometimes called a 'parking orbit'. Relay satellites can be placed in parking orbits, so that television programmes can be transmitted continuously from one part of the world to another. *Syncom* was a satellite used for transmission of the Tokio Olympic Games in 1964.

Since $T = 24$ hours, the radius R can be found from (i). Thus from

$$R = \sqrt[3]{\frac{T^2 g r^2}{4\pi^2}} \quad \text{and} \quad g = 9.8 \text{ m s}^{-2}, r = 6.4 \times 10^6 \text{ m},$$

$$\therefore R = \sqrt[3]{\frac{(24 \times 3600)^2 \times 9.8 \times (6.4 \times 10^6)^2}{4\pi^2}} = 42400 \text{ km}$$

The height above the earth's surface of the parking orbit

$$= R - r = 42\,400 - 6\,400 = 36\,000 \text{ km}.$$

In the orbit, the velocity of the satellite

$$= \frac{2\pi R}{T} = \frac{2\pi \times 42\,400}{24 \times 3600 \text{ seconds}} = 3.1 \text{ km s}^{-1}.$$

Weightlessness

When a rocket is fired to launch a spacecraft and astronaut into orbit round the earth, the initial acceleration must be very high owing to the large initial thrust required. This acceleration, a , is of the order of $15g$, where g is the gravitational acceleration at the earth's surface.

Suppose S is the reaction of the couch to which the astronaut is initially strapped. Fig. 2.23 (i). Then, from $F = ma$, $S - mg = ma = m \cdot 15g$, where m is the mass of the astronaut. Thus $S = 16mg$. This force is 16 times the weight of the astronaut and thus, initially, he experiences a large force.

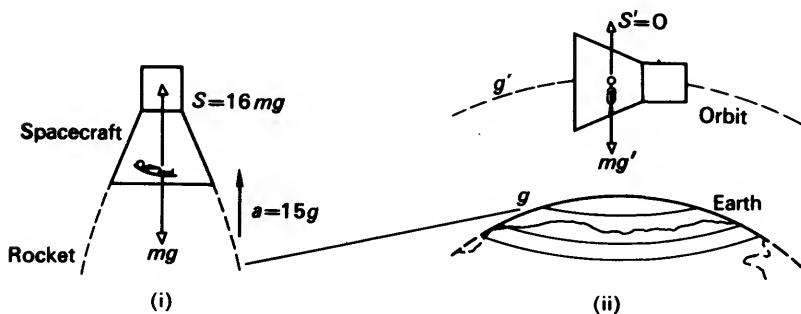


FIG. 2.23 Weight and weightlessness

In orbit, however, the state of affairs is different. This time the acceleration of the spacecraft and astronaut are both g' in magnitude,

where g' is the acceleration due to gravity outside the spacecraft at the particular height of the orbit. Fig. 2.23 (ii). If S' is the reaction of the surface of the spacecraft in contact with the astronaut, then, for circular motion,

$$F = mg' - S' = ma = mg'.$$

Thus $S' = 0$. Consequently the astronaut becomes 'weightless'; he experiences no reaction at the floor when he walks about, for example. At the earth's surface we feel the reaction at the ground and are thus conscious of our weight. Inside a lift which is falling fast, the reaction at our feet diminishes. If the lift falls freely, the acceleration of objects inside is the same as that outside and hence the reaction on them is zero. This produces the sensation of 'weightlessness'. In orbit, as in Fig. 2.23 (ii), objects inside a spacecraft are also in 'free fall' because they have the same acceleration g' as the spacecraft. Consequently the sensation of weightlessness is experienced.

EXAMPLE

A satellite is to be put into orbit 500 km above the earth's surface. If its vertical velocity after launching is 2000 m s^{-1} at this height, calculate the magnitude and direction of the impulse required to put the satellite directly into orbit, if its mass is 50 kg. Assume $g = 10 \text{ m s}^{-2}$; radius of earth, $R = 6400 \text{ km}$.

Suppose u is the velocity required for orbit, radius r . Then, with usual notation,

$$\frac{mu^2}{r} = \frac{GmM}{r^2} = \frac{gR^2m}{r^2}, \text{ as } \frac{GM}{R^2} = g.$$

$$\therefore u^2 = \frac{gR^2}{r}.$$

Now $R = 6400 \text{ km}$, $r = 6900 \text{ km}$, $g = 10 \text{ m s}^{-2}$.

$$\therefore u^2 = \frac{10 \times (6400 \times 10^3)^2}{6900 \times 10^3}.$$

$$\therefore u = 7700 \text{ m s}^{-1} \text{ (approx.)}.$$

At this height, vertical momentum

$$U_Y = mv = 50 \times 2000 = 100\,000 \text{ kg m s}^{-1}.$$

Fig. 2.24.

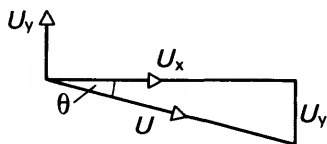


FIG. 2.24 Example

Horizontal momentum required $U_X = mu = 50 \times 7700 = 385\,000 \text{ kg m s}^{-1}$.

$$\therefore \text{impulse needed, } U = \sqrt{U_Y^2 + U_X^2} = \sqrt{100\,000^2 + 385\,000^2} \\ = 4.0 \times 10^5 \text{ kg m s}^{-1} \quad (1)$$

Direction. The angle θ made by the total impulse with the horizontal or orbit tangent is given by $\tan \theta = U_Y/U_X = 100\,000/385\,000 = 0.260$. Thus $\theta = 14.6^\circ$.

Magnitudes of acceleration due to gravity

(i) *Above the earth's surface.* Consider an object of mass m in an orbit of radius R from the centre, where $R > r$, the radius of the earth. Then, if g' is the acceleration due to gravity at this place,

$$mg' = \frac{GmM}{R^2} \quad (i)$$

But, if g is the acceleration due to gravity at the earth's surface,

$$mg = \frac{GmM}{r^2} \quad \text{. (ii)}$$

Dividing (i) by (ii), $\therefore \frac{g'}{g} = \frac{r^2}{R^2}$, or $g' = \frac{r^2}{R^2} \cdot g$.

Thus above the earth's surface, the acceleration due to gravity g' varies *inversely as the square of the distance* from the centre. Fig. 2.25.

For a height h above the earth, $R = r + h$.

$$\begin{aligned} \therefore g' &= \frac{r^2}{(r+h)^2} \cdot g = \frac{1}{\left(1 + \frac{h}{r}\right)^2} \cdot g \\ &= \left(1 + \frac{h}{r}\right)^{-2} \cdot g = \left(1 - \frac{2h}{r}\right)g, \end{aligned}$$

since powers of $(h/r)^2$ and higher can be neglected when h is small compared with r .

$\therefore g - g' =$ reduction in acceleration due to gravity.

$$= \frac{2h}{r} \cdot g \quad \text{. (1)}$$

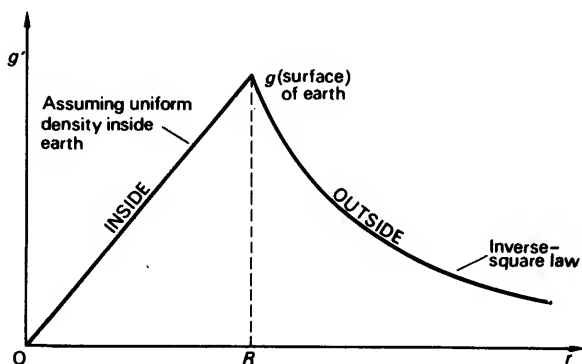


FIG. 2.25 Variation of g

(ii) *Below the earth's surface.* Consider an object of mass m at a point below the earth's surface. If its distance from the centre is b , the 'effective' mass M' of the earth which attracts it is that contained in a sphere of radius b . Assuming a constant density, then, since the mass of a sphere is proportional to *radius*³,

$$M' = \frac{b^3}{r^3} M,$$

Now

$$GM/r^2 = g.$$

$$\therefore v = \sqrt{2gr}.$$

$$\therefore v = \sqrt{2 \times 9.8 \times 6.4 \times 10^6} = 11 \times 10^3 \text{ m s}^{-1} = 11 \text{ km s}^{-1} \text{ (approx).}$$

With an initial velocity, then, of about 11 km s^{-1} , a rocket will completely escape from the gravitational attraction of the earth. It can be made to travel towards the moon, for example, so that eventually it comes under the gravitational attraction of this planet. At present, 'soft' landings on the moon have been made by firing retarding retro rockets.

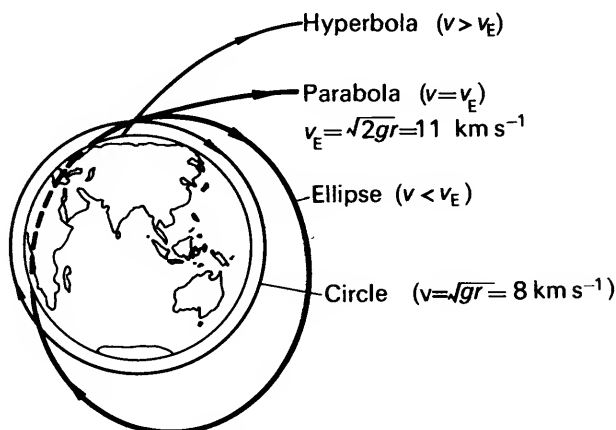


FIG. 2.26 Orbits

Summarising, with a velocity of about 8 km s^{-1} , a satellite can describe a circular orbit close to the earth's surface (p. 64). With a velocity greater than 8 km s^{-1} but less than 11 km s^{-1} , a satellite describes an elliptical orbit round the earth. Its maximum and minimum height in the orbit depends on its particular velocity. Fig. 2.26 illustrates the possible orbits of a satellite launched from the earth.

The molecules of air at normal temperatures and pressures have an average velocity of the order of 480 m s^{-1} or 0.48 km s^{-1} which is much less than the velocity of escape. Many molecules move with higher velocity than 0.48 km s^{-1} but gravitational attraction keeps the atmosphere round the earth. The gravitational attraction of the moon is much less than that of the earth and this accounts for the lack of atmosphere round the moon.

EXERCISES 2

(Assume $g = 10 \text{ m s}^{-2}$)

What are the missing words in the statements 1-6?

1. The force towards the centre in circular motion is called the ... force.
2. In simple harmonic motion, the maximum kinetic energy occurs at the ... of the oscillation.

3. The constant of gravitation G is related to g by ...
4. In simple harmonic motion, the maximum potential energy occurs at the ... of the oscillation.
5. Outside the earth, the acceleration due to gravity is proportional to ... from the centre.
6. A satellite in orbit in an equatorial plane round the earth will stay at the same place above the earth if its period is ... hours.

Which of the following answers, A, B, C, D or E, do you consider is the correct one in the statements 7–10?

7. The earth retains its atmosphere because *A* the earth is spherical, *B* the velocity of escape is greater than the mean speed of molecules, *C* the constant of gravitation is a universal constant, *D* the velocity of escape is less than the mean speed of molecules, *E* gases are lighter than solids.
8. In simple harmonic motion, the moving object has *A* only kinetic energy, *B* mean kinetic energy greater than the mean potential energy, *C* total energy equal to the sum of the maximum kinetic energy and maximum potential energy, *D* mean kinetic energy equal to the mean potential energy, *E* minimum potential energy at the centre of oscillation.
9. If r is the radius of the earth and g is the acceleration at its surface, then the acceleration g' at an orbit distance R from the centre of the earth is given by *A* $g'/g = r/R$, *B* $g'/g = r^2/R^2$, *C* $g'/g = R^2/r^2$, *D* $g'/g = (R-r)^2/r^2$, *E* $g'/g = (R-r)/r$.

10. When water in a bucket is whirled fast overhead, the water does not fall out at the top of the motion because *A* the centripetal force on the water is greater than the weight of water, *B* the force on the water is opposite to gravity, *C* the reaction of the bucket on the water is zero, *D* the centripetal force on the water is less than the weight of water, *E* atmospheric pressure counteracts the weight.

Circular Motion

11. An object of mass 4 kg moves round a circle of radius 6 m with a constant speed of 12 m s^{-1} . Calculate (i) the angular velocity, (ii) the force towards the centre.
12. An object of mass 10 kg is whirled round a horizontal circle of radius 4 m by a revolving string inclined to the vertical. If the uniform speed of the object is 5 m s^{-1} , calculate (i) the tension in the string in kgf, (ii) the angle of inclination of the string to the vertical.
13. A racing-car of 1000 kg moves round a banked track at a constant speed of 108 km h^{-1} . Assuming the total reaction at the wheels is normal to the track, and the horizontal radius of the track is 100 m, calculate the angle of inclination of the track to the horizontal and the reaction at the wheels.
14. An object of mass 8.0 kg is whirled round in a vertical circle of radius 2 m with a constant velocity of 6 m s^{-1} . Calculate the maximum and minimum tensions in the string.

15. Define the terms (a) *acceleration*, and (b) *force*. Show that the acceleration of a body moving in a circular path of radius r with uniform speed v is v^2/r , and draw a diagram to show the direction of the acceleration.

A small body of mass m is attached to one end of a light inelastic string of

length l . The other end of the string is fixed. The string is initially held taut and horizontal, and the body is then released. Find the values of the following quantities when the string reaches the vertical position: (a) the kinetic energy of the body, (b) the velocity of the body, (c) the acceleration of the body, and (d) the tension in the string. (O. & C.)

16. Explain what is meant by *angular velocity*. Derive an expression for the force required to make a particle of mass m move in a circle of radius r with uniform angular velocity ω .

A stone of mass 500 g is attached to a string of length 50 cm which will break if the tension in it exceeds 2.0 kgf. The stone is whirled in a vertical circle, the axis of rotation being at a height of 100 cm above the ground. The angular speed is very slowly increased until the string breaks. In what position is this break most likely to occur, and at what angular speed? Where will the stone hit the ground? (C.)

Simple Harmonic Motion

17. An object moving with simple harmonic motion has an amplitude of 2 cm and a frequency of 20 Hz. Calculate (i) the period of oscillation, (ii) the acceleration at the middle and end of an oscillation, (iii) the velocities at the corresponding instants.

18. Calculate the length in centimetres of a simple pendulum which has a period of 2 seconds. If the amplitude of swing is 2 cm, calculate the velocity and acceleration of the bob (i) at the end of a swing, (ii) at the middle, (iii) 1 cm from the centre of oscillation.

19. Define *simple harmonic motion*. An elastic string is extended 1 cm when a small weight is attached at the lower end. If the weight is pulled down $\frac{1}{4}$ cm and then released, show that it moves with simple harmonic motion, and find the period.

20. A uniform wooden rod floats upright in water with a length of 30 cm immersed. If the rod is depressed slightly and then released, prove that its motion is simple harmonic and calculate the period.

21. A simple pendulum, has a period of 4.2 seconds. When the pendulum is shortened by 1 m, the period is 3.7 seconds. From these measurements, calculate the acceleration due to gravity and the original length of the pendulum.

22. What is *simple harmonic motion*? Show how it is related to the uniform motion of a particle with velocity v in a circle of radius r .

A steel strip, clamped at one end, vibrates with a frequency of 50 Hz and an amplitude of 8 mm at the free end. Find (a) the velocity of the end when passing through the zero position, (b) the acceleration at the maximum displacement.

23. Explain what is meant by *simple harmonic motion*.

Show that the vertical oscillations of a mass suspended by a light helical spring are simple harmonic and describe an experiment with the spring to determine the acceleration due to gravity.

A small mass rests on a horizontal platform which vibrates vertically in simple harmonic motion with a period of 0.50 second. Find the maximum amplitude of the motion which will allow the mass to remain in contact with the platform throughout the motion. (L.)

24. Define simple harmonic motion and state a formula for its period. Show that under suitable conditions the motion of a simple pendulum is simple harmonic and hence obtain an expression for its period.

If a pendulum bob is suspended from an inaccessible point, by a string whose length may be varied, describe how to determine (a) the acceleration due to gravity, (b) the height of the point of suspension above the floor.

How and why does the value of the acceleration due to gravity at the poles differ from its value at the equator? (L.)

25. Derive an expression for the time period of vertical oscillations of small amplitude of a mass suspended from the free end of a light helical spring.

What deformation of the wire of the spring occurs when the mass moves? (N.)

26. Give two practical examples of oscillatory motion which approximate to simple harmonic motion. What conditions must be satisfied if the approximations are to be good ones.

A point mass moves with simple harmonic motion. Draw on the same axes sketch graphs to show the variation with position of (a) the potential energy, (b) the kinetic energy, and (c) the total energy of the particle.

A particle rests on a horizontal platform which is moving vertically in simple harmonic motion with an amplitude of 10 cm. Above a certain frequency, the thrust between the particle and the platform would become zero at some point in the motion. What is this frequency, and at what point in the motion does the thrust become zero at this frequency? (C.)

27. In what circumstances will a particle execute simple harmonic motion? Show how simple harmonic motion can be considered to be the projection on the diameter of a circle of the motion of a particle describing the circle with uniform speed.

The balance wheel of a watch vibrates with an angular amplitude of π radians and a period of 0.5 second. Calculate (a) the maximum angular speed, (b) the angular speed when the displacement is $\pi/2$, and (c) the angular acceleration when the displacement is $\pi/4$. If the radius of the wheel is r , calculate the maximum radial force acting on a small dust particle of mass m situated on the rim of the wheel. (O. & C.)

28. Prove that the bob of a simple pendulum may move with simple harmonic motion, and find an expression for its period.

Describe with full details how you would perform an experiment, based on the expression derived, to measure the value of the acceleration due to gravity. What factors would influence your choice of (a) the length of the pendulum, (b) the material of the bob, and (c) the number of swings to be timed? (O. & C.)

29. Define *simple harmonic motion* and show that the free oscillations of a simple pendulum are simple harmonic for small amplitudes.

Explain what is meant by damping of oscillations and describe an experiment to illustrate the effects of damping on the motion of a simple pendulum. Briefly discuss the difficulties you would encounter and indicate qualitatively the results you would expect to observe. (O. & C.)

30. What is meant by simple harmonic motion? Obtain an expression for the kinetic energy of a body of mass m , which is performing S.H.M. of amplitude a and period $2\pi/\omega$, when its displacement from the origin is x .

Describe an experiment, or experiments, to verify that a mass oscillating at the end of a helical spring moves with simple harmonic motion. (C.)

31. State the dynamical condition under which a particle will describe simple harmonic motion. Show that it is approximately fulfilled in the case of the bob of a simple pendulum, and derive, from first principles, an expression for the period of the pendulum.

Explain how it can be demonstrated from observations on simple pendulums, that the weight of a body at a given place is proportional to its mass. (O. & C.)

32. Define *simple harmonic motion*. Show that a heavy body supported by a light spiral spring executes simple harmonic motion when displaced vertically from its equilibrium position by an amount which does not exceed a certain value and then released. How would you determine experimentally the maximum amplitude for simple harmonic motion?

A spiral spring gives a displacement of 5 cm for a load of 500 g. Find the maximum displacement produced when a mass of 80 g is dropped from a height of 10 cm on to a light pan attached to the spring. (N.)

Gravitation

33. Calculate the force of attraction between two small objects of mass 5 and 8 kg respectively which are 10 cm apart. ($G = 6.7 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$.)

34. If the acceleration due to gravity is 9.8 m s^{-2} and the radius of the earth is 6400 km, calculate a value for the mass of the earth. ($G = 6.7 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$.) Give the theory.

35. Assuming that the mean density of the earth is 5500 kg m^{-3} , that the constant of gravitation is $6.7 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$, and that the radius of the earth is 6400 km, find a value for the acceleration due to gravity at the earth's surface. Derive the formula used.

36. How do you account for the sensation of 'weightlessness' experienced by the occupant of a space capsule (a) in a circular orbit round the earth, (b) in outer space? Give one other instance in which an object would be 'weightless'. (N.)

37. State Newton's law of universal gravitation. Distinguish between the gravitational constant (G) and the acceleration due to gravity (g) and show the relation between them.

Describe an experiment by which the value of g may be determined. Indicate the measurements taken and how to calculate the result. Derive any formula used. (L.)

38. State *Newton's law of gravitation*. What experimental evidence is there for the validity of this law?

A binary star consists of two dense spherical masses of 10^{30} kg and $2 \times 10^{30} \text{ kg}$ whose centres are 10^7 km apart and which rotate together with a uniform angular velocity ω about an axis which intersects the line joining their centres. Assuming that the only forces acting on the stars arise from their mutual gravitational attraction and that each mass may be taken to act at its centre, show that the axis of rotation passes through the centre of mass of the system and find the value of ω . ($G = 6.7 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$.) (O. & C.)

39. Assuming that the planets are moving in circular orbits, apply Kepler's laws to show that the acceleration of a planet is inversely proportional to the square of its distance from the sun. Explain the significance of this and show clearly how it leads to Newton's law of universal gravitation.

Obtain the value of g from the motion of the moon, assuming that its period of rotation round the earth is 27 days 8 hours and that the radius of its orbit is 60.1 times the radius of the earth. (Radius of earth = $6.36 \times 10^6 \text{ m}$.) (N.)

40. Explain what is meant by the *gravitation constant* (G), and describe an accurate laboratory method of measuring it. Give an outline of the theory of your method.

Assuming that the earth is a sphere of radius 6370 km and that $G = 6.66 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$, calculate the mean density of the earth. (O. & C.)

41. Assuming the earth to be perfectly spherical, give sketch graphs to show how (a) the acceleration due to gravity, (b) the gravitational potential due to the earth's mass, vary with distance from the surface of the earth for points external to it. If any other assumption has been made, state what it is.

Explain why, even if the earth were a perfect sphere, the period of oscillation of a simple pendulum at the poles would not be the same as at the equator. Still assuming the earth to be perfectly spherical, discuss whether the velocity required to project a body vertically upwards, so that it rises to a given height, depends on the position on the earth from which it is projected. (C.)

42. Explain what is meant by the *constant of gravitation*. Describe a laboratory experiment to determine it, showing how the result is obtained from the observations.

A proposed communication satellite would revolve round the earth in a circular orbit in the equatorial plane, at a height of 35880 km above the earth's surface. Find the period of revolution of the satellite in hours, and comment on the result. (Radius of earth = 6370 km, mass of earth = $5.98 \times 10^{24} \text{ kg}$ constant of gravitation = $6.66 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$.) (N.)

chapter three

Rotation of Rigid Bodies

So far in this book we have considered the equations of motion and other dynamical formulae associated with a particle. In practice, however, an object is made of millions of particles, each at different places, and we need now to consider the motion of moving objects.

Moment of Inertia, I

Suppose a rigid object is rotating about a fixed axis O , and a particle A of the object makes an angle θ with a fixed line OY in space at some instant, Fig. 3.1. The angular velocity, $d\theta/dt$ or ω , of every particle about O is the same, since we are dealing with a rigid body, and the velocity v_1 of A at this instant is given by $r_1\omega$, where $r_1 = OA$. Thus the kinetic energy of $A = \frac{1}{2}m_1v_1^2 = \frac{1}{2}m_1r_1^2\omega^2$. Similarly, the kinetic energy of another particle of the body $= \frac{1}{2}m_2r_2^2\omega^2$, where r_2 is its distance from O and m_2 is its mass. In this way we see that the kinetic energy, K.E., of the whole object is given by

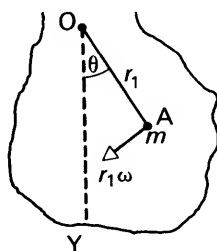


FIG. 3.1
Rotating rigid body

$$\begin{aligned} \text{K.E.} &= \frac{1}{2}m_1r_1^2\omega^2 + \frac{1}{2}m_2r_2^2\omega^2 + \frac{1}{2}m_3r_3^2\omega^2 + \dots \\ &= \frac{1}{2}\omega^2(m_1r_1^2 + m_2r_2^2 + m_3r_3^2 + \dots) \\ &= \frac{1}{2}\omega^2(\Sigma mr^2), \end{aligned}$$

where Σmr^2 represents the sum of the magnitudes of ' mr^2 ' for all the particles of the object. We shall see shortly how the quantity Σmr^2 can be calculated for a particular object. The magnitude of Σmr^2 is known as the *moment of inertia* of the object about the axis concerned, and we shall denote it by the symbol I . Thus

$$\text{Kinetic energy, K.E.,} = \frac{1}{2}I\omega^2. \quad (1)$$

The units of I are kg metre^2 (kg m^2). The unit of ω is ' radian s^{-1} ' (rad s^{-1}). Thus if $I = 2 \text{ kg m}^2$ and $\omega = 3 \text{ rad s}^{-1}$, then

$$\text{K.E.} = \frac{1}{2}I\omega^2 = \frac{1}{2} \times 2 \times 3^2 \text{ joule} = 9 \text{ J.}$$

The kinetic energy of a particle of mass m moving with a velocity v is $\frac{1}{2}mv^2$. It will thus be noted that the formula for the kinetic energy of a rotating object is similar to that of a moving particle, the mass m being replaced by the moment of inertia I and the velocity v being replaced by the angular velocity ω . As we shall require values of I , the moment of inertia of several objects about a particular axis will first be calculated.

Moment of Inertia of Uniform Rod

(1) *About axis through middle.* The moment of inertia of a small element δx about an axis PQ through its centre O perpendicular to the length $= \left(\frac{\delta x}{l}M\right)x^2$, where l is the length of the rod, M is its mass, and x is the distance of the small element from O, Fig. 3.2.

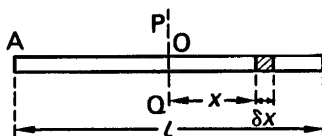


FIG. 3.2 Moment of inertia—uniform rod

$$\begin{aligned}\therefore \text{moment of inertia, } I, &= 2 \int_0^{l/2} \left(\frac{dx}{l}M\right)x^2 \\ &= \frac{2M}{l} \int_0^{l/2} x^2 dx = \frac{Ml^2}{12}\end{aligned}\quad (1)$$

Thus if the mass of the rod is 60 g and its length is 20 cm, $M = 6 \times 10^{-2}$ kg, $l = 0.2$ m, and $I = 6 \times 10^{-2} \times 0.2^2/12 = 2 \times 10^{-4}$ kg m².

(2) *About the axis through one end, A.* In this case, measuring distances x from A instead of O,

$$\text{moment of inertia, } I, = \int_0^l \left(\frac{dx}{l}M\right) \times x^2 = \frac{Ml^2}{3}\quad (2)$$

Moment of Inertia of Ring

Every element of the ring is the same distance from the centre. Hence the moment of inertia about an axis through the centre perpendicular to the plane of the ring $= Ma^2$, where M is the mass of the ring and a is its radius.

Moment of Inertia of Circular Disc

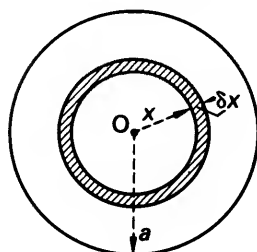


FIG. 3.3
Moment of inertia—disc

Consider the moment of inertia of a circular disc about an axis through its centre perpendicular to its plane, Fig. 3.3. If we take a small ring of the disc enclosed between radii x and $x + \delta x$, its mass $= \frac{2\pi x \delta x}{\pi a^2}M$, where a is the radius of the disc and M is its mass. Each element of the ring is distant x from the centre, and hence the moment of inertia of the ring about the axis through O $= \left(\frac{2\pi x \delta x}{\pi a^2}M\right) \times x^2$

$$\begin{aligned}\therefore \text{moment of inertia of whole disc} &= \int_0^a \frac{2\pi x dx}{\pi a^2} M \times x^2 \\ &= \frac{Ma^2}{2}\end{aligned}\quad (1)$$

Thus if the disc weighs 60 g and has a radius of 10 cm, $M = 60 \text{ g} = 6 \times 10^{-2} \text{ kg}$, $a = 0.1 \text{ m}$, so that $I = 6 \times 10^{-2} \times 0.1^2/2 = 3 \times 10^{-4} \text{ kg m}^2$.

Moment of Inertia of Cylinder

If a cylinder is *solid*, its moment of inertia about the axis of symmetry is the sum of the moments of inertia of discs into which we may imagine the cylinder cut. The moment of inertia of each disc $= \frac{1}{2} \text{ mass} \times a^2$, where a is the radius; and hence, if M is the mass of the cylinder,

$$\text{moment of inertia of solid cylinder} = \frac{1}{2}Ma^2 \quad (i)$$

If a cylinder is *hollow*, its moment of inertia about the axis of symmetry is the sum of the moments of inertia of the curved surface and that of the two ends, assuming the cylinder is closed at both ends. Suppose a is the radius, h is the height of the cylinder, and σ is the mass per unit area of the surface. Then

$$\text{mass of curved surface} = 2\pi ah\sigma,$$

$$\text{and moment of inertia about axis} = \text{mass} \times a^2 = 2\pi a^3h\sigma,$$

since we can imagine the surface cut into rings.

The moment of inertia of one end of the cylinder $= \text{mass} \times a^2/2 = \pi a^2\sigma \times a^2/2 = \pi a^4\sigma/2$. Hence the moment of inertia of both ends $= \pi a^4\sigma$.

$$\therefore \text{moment inertia of cylinder, } I, = 2\pi a^3h\sigma + \pi a^4\sigma.$$

$$\text{The mass of the cylinder, } M, = 2\pi ah\sigma + 2\pi a^2\sigma$$

$$\begin{aligned}\therefore I &= \frac{2\pi a^3h\sigma + \pi a^4\sigma}{2\pi ah\sigma + 2\pi a^2\sigma} M \\ &= \frac{2a^2h + a^3}{2h + 2a} M \\ &= \frac{1}{2}Ma^2 + \frac{a^2h}{2h + 2a} M\end{aligned}\quad (ii)$$

If a hollow and a solid cylinder have the same mass M and the same radius and height, it can be seen from (i) and (ii) that the moment of inertia of the hollow cylinder is greater than that of the solid cylinder about the axis of symmetry. This is because the mass is distributed on the average at a greater distance from the axis in the former case.

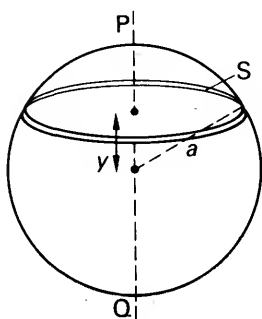


FIG. 3.4
Moment of inertia—sphere

Moment of Inertia of Sphere

The moment of inertia of a sphere about an axis PQ through its centre can be found by cutting thin discs such as S perpendicular to the axis, Fig. 3.4. The volume of the disc, of thickness δy and distance y from the centre,

$$= \pi r^2 \delta y = \pi(a^2 - y^2)\delta y.$$

$$\therefore \text{mass } M' \text{ of disc} = \frac{\pi(a^2 - y^2)\delta y}{4\pi a^3/3} M$$

$$= \frac{3M}{4a^3}(a^2 - y^2)\delta y,$$

where M is the mass of the sphere and a is its radius, since the volume of the sphere $= 4\pi a^3/3$. Now the moment of inertia of the disc about PQ

$$= M' \times \frac{\text{radius}^2}{2} = \frac{3M}{4a^3}(a^2 - y^2)\delta y \times \frac{(a^2 - y^2)}{2}$$

$$\therefore \text{moment of inertia of sphere} = \frac{3M}{8a^3} \int_{-a}^{+a} (a^4 - 2a^2 y^2 + y^4) dy$$

$$= \frac{2}{5} Ma^2 \quad (1)$$

Thus if the sphere has mass 4 kg and a radius of 0.2 m, the moment of inertia $= \frac{2}{5} \times 4 \times 0.2^2 = 0.064 \text{ kg m}^2$.

Radius of Gyration

The moment of inertia of an object about an axis, Σmr^2 , is sometimes written as Mk^2 , where M is the mass of the object and k is a quantity called the *radius of gyration* about the axis. For example, the moment of inertia of a rod about an axis through one end $= Ml^2/3$ (p. 76) $= M(l/\sqrt{3})^2$. Thus the radius of gyration, $k, = l/\sqrt{3} = 0.58l$. The moment of inertia of a sphere about its centre $= \frac{2}{5}Ma^2 = M \times (\sqrt{\frac{2}{5}}a)^2$. Thus the radius of gyration, $k, = \sqrt{\frac{2}{5}}a = 0.63a$ in this case.

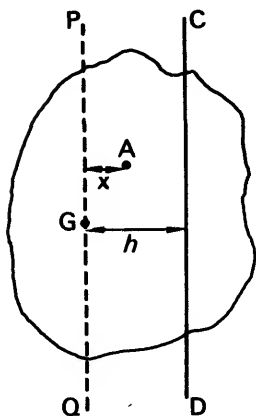


FIG. 3.5
Theorem of parallel axes

Relation Between Moment of Inertia About C.G. and Parallel Axis.

Suppose I is the moment of inertia of a body about an axis CD and I_G is the moment of inertia about a parallel axis PQ through the centre of gravity, G, distant h from the axis CD, Fig. 3.5. If A is a particle of mass m whose distance from PQ is x , its moment of inertia about CD $= m(h-x)^2$

$$\therefore I = \Sigma m(h-x)^2 = \Sigma mh^2 + \Sigma mx^2 - \Sigma 2mhx.$$

Now $\Sigma mh^2 = h^2 \times \Sigma m = Mh^2$, where M is the total mass of the object, and $\Sigma mx^2 = I_G$, the moment of inertia through the centre of gravity.

Also, $\Sigma 2mhx = 2h\Sigma mx = 0$,

since Σmx , the sum of the moments about the centre of gravity, is zero; this follows because the moment of the resultant (the weight) about G is zero.

$$\therefore I = I_G + Mh^2 \quad (1)$$

From this result, it follows that the moment of inertia, I , of a disc of radius a and mass M about an axis through a point on its circumference $= I_G + Ma^2$, since $h = a =$ radius of disc in this case. But $I_G =$ moment of inertia about the centre $= Ma^2/2$ (p.77).

$$\therefore \text{moment of inertia, } I, = \frac{Ma^2}{2} + Ma^2 = \frac{3Ma^2}{2}.$$

Similarly the moment of inertia of a sphere of radius a and mass M about an axis through a point on its circumference $= I_G + Ma^2 = 2Ma^2/5 + Ma^2 = 7Ma^2/5$, since I_G , the moment of inertia about an axis through its centre, is $2Ma^2/5$.

Relation Between Moments of Inertia about Perpendicular Axes

Suppose OX, OY are any two perpendicular axes and OZ is an axis perpendicular to OX and OY , Fig. 3.6 (i). The moment of inertia, I , of

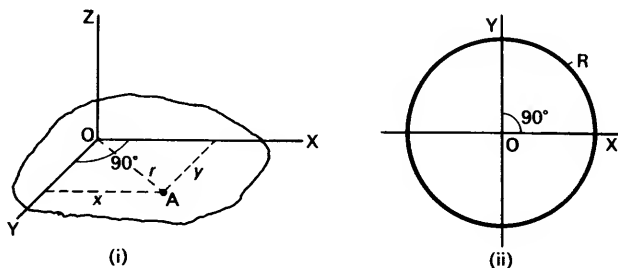


FIG. 3.6 Theorem of perpendicular axes

a body about the axis $OZ = \Sigma mr^2$, where r is the distance of a particle A from OZ and m is its mass. But $r^2 = x^2 + y^2$, where x, y are the distances of A from the axis OY, OX respectively.

$$\therefore I = \Sigma m(x^2 + y^2) = \Sigma mx^2 + \Sigma my^2.$$

$$\therefore I = I_y + I_x \quad (1)$$

where I_y, I_x are the moments of inertia about OX, OY respectively.

As a simple application, consider a ring R and two perpendicular axes OX, OY in its plane, Fig. 3.6 (ii). Then from the above result,

$I_y + I_x = I = \text{moment of inertia through } O \text{ perpendicular to ring.}$

$$\therefore I_y + I_x = Ma^2.$$

But $I_y = I_x$ by symmetry.

$$\therefore I_x + I_x = Ma^2,$$

$$\therefore I_x = \frac{Ma^2}{2}.$$

This is the moment of inertia of the ring about any diameter in its plane.

In the same way, the moment of inertia, I , of a *disc* about a diameter in its plane is given by

$$I + I = \frac{Ma^2}{2},$$

since the moments of inertia, I , about the two perpendicular diameters are the same and $Ma^2/2$ is the moment of inertia of the disc about an axis perpendicular to its plane.

$$\therefore I = \frac{Ma^2}{4}.$$

Couple on a Rigid Body

Consider a rigid body rotating about a fixed axis O with an angular velocity ω at some instant. Fig. 3.7.

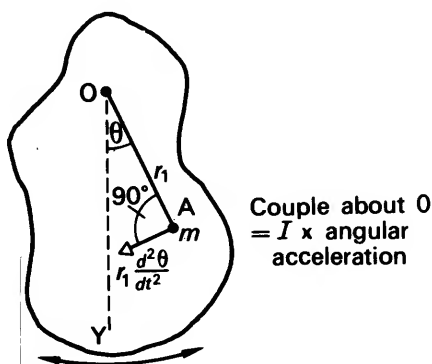


FIG. 3.7 Couple on rigid body

The force acting on the particle $A = m_1 \times \text{acceleration} = m_1 \times \frac{d}{dt}(r_1 \omega) = m_1 \times r_1 \frac{d\omega}{dt} = m_1 r_1 \frac{d^2 \theta}{dt^2}$, since $\omega = \frac{d\theta}{dt}$. The moment of this

force about the axis O = force \times perpendicular distance from O = $m_1 r_1 \frac{d^2\theta}{dt^2} \times r_1$, since the force acts perpendicularly to the line OA .

$$\therefore \text{moment or torque} = m_1 r_1^2 \frac{d^2\theta}{dt^2}.$$

\therefore total moment of all forces on body about O , or *torque*,

$$\begin{aligned} &= m_1 r_1^2 \frac{d^2\theta}{dt^2} + m_2 r_2^2 \frac{d^2\theta}{dt^2} + m_3 r_3^2 \frac{d^2\theta}{dt^2} + \dots \\ &= (\Sigma m r^2) \times \frac{d^2\theta}{dt^2}, \end{aligned}$$

since the angular acceleration, $d^2\theta/dt^2$, about O is the same for all particles.

$$\therefore \text{total torque about } O = I \frac{d^2\theta}{dt^2}, \quad (1)$$

where $I = \Sigma m r^2$ = moment of inertia about O . The moment about O is produced by external forces which together act as a *couple* of torque C say. Thus, for any rotating rigid body,

$$\text{Couple, } C = I \frac{d^2\theta}{dt^2}.$$

This result is analogous to the case of a particle of mass m which undergoes an acceleration a when a force F acts on it. Here $F = ma$. In place of F we have a *couple* C for a rotating rigid object; in place of m we have the *moment of inertia* I ; and in place of linear acceleration a , we have *angular acceleration* $d^2\theta/dt^2 (d\omega/dt)$.

EXAMPLES

1. A heavy flywheel of mass 15 kg and radius 20 cm is mounted on a horizontal axle of radius 1 cm and negligible mass compared with the flywheel. Neglecting friction, find (i) the angular acceleration if a force of 4 kgf is applied tangentially to the axle, (ii) the angular velocity of the flywheel after 10 seconds.

$$(i) \quad \text{Moment of inertia} = \frac{Ma^2}{2} = \frac{15 \times 0.2^2}{2} = 0.3 \text{ kg m}^2.$$

$$\text{Couple } C = 4 \times 9.8 \text{ (N)} \times 0.01 \text{ (m)} = 0.4 \text{ N m approx.}$$

$$\therefore \text{angular acceleration} = \frac{0.4}{0.3} = 1.3 \text{ rad s}^{-2}.$$

(ii) After 10 seconds, angular velocity = angular acceleration \times time.

$$= 1.3 \times 10 = 13 \text{ rad s}^{-1}.$$

2. The moment of inertia of a solid flywheel about its axis is 0.1 kg m². It is set in rotation by applying a tangential force of 2 kgf with a rope wound round the circumference, the radius of the wheel being 10 cm. Calculate the

angular acceleration of the flywheel. What would be the acceleration if a mass of 2 kg were hung from the end of the rope? (O. & C.)

Couple $C = I \frac{d^2\theta}{dt^2}$ = moment of inertia \times angular acceleration.

Now $C = 2 \times 9.8 \times 0.1$ N m.

$$\therefore \text{angular acceleration} = \frac{2 \times 9.8 \times 0.1}{0.1} \\ = 19.6 \text{ rad s}^{-2}.$$

If a mass of 2 kg is hung from the end of the rope, it moves down with an acceleration a . Fig. 3.8. In this case, if T is the tension in the rope,

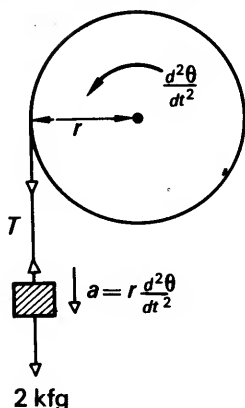


FIG. 3.8 Example

$$mg - T = ma \quad (1)$$

For the flywheel, $T.r = \text{couple} = I \frac{d^2\theta}{dt^2} \quad (2)$

where r is the radius of the flywheel. Now the mass of 2 kg descends a distance given by $r\theta$, where θ is the angle the flywheel has turned. Hence the acceleration $a = r d^2\theta/dt^2$. Substituting in (1),

$$\therefore mg - T = mr \frac{d^2\theta}{dt^2}.$$

$$\therefore mgr - T.r = mr^2 \frac{d^2\theta}{dt^2} \quad (3)$$

Adding (2) and (3),

$$\therefore mgr = (I + mr^2) \frac{d^2\theta}{dt^2}.$$

$$\therefore \frac{d^2\theta}{dt^2} = \frac{mgr}{I + mr^2} = \frac{2 \times 10 \times 0.1}{0.1 + 2 \times 0.1^2} \\ = 16.7 \text{ rad s}^{-2}.$$

using $g = 10 \text{ m s}^{-2}$.

Angular Momentum and Conservation

In linear or straight-line motion, an important property of a moving object is its linear momentum (p. 18). When an object spins or rotates about an axis, its *angular momentum* plays an important part in its motion.

Consider a particle A of a rigid object rotating about an axis O. Fig. 3.9. The momentum of A = mass \times velocity = $m_1 v = m_1 r_1 \omega$. The 'angular momentum' of A about O is defined as the *moment of the momentum* about O. Its magnitude is thus $m_1 v \times p$, where p is the perpendicular distance from O to the direction of v . Thus angular momentum of A = $m_1 v p = m_1 r_1 \omega \times r_1 = m_1 r_1^2 \omega$.

$$\therefore \text{total angular momentum of whole body} = \Sigma m_1 r_1^2 \omega = \omega \Sigma m_1 r_1^2 = I\omega,$$

where I is the moment of inertia of the body about O .

Angular momentum is analogous to 'linear momentum', mv , in the dynamics of a moving particle. In place of m we have I , the moment of inertia; in place of v we have ω , the angular velocity.

Further, the *conservation of angular momentum*, which corresponds to the conservation of linear momentum, states that *the angular momentum about an axis of a given rotating body or system of bodies is constant, if no external couple acts about that axis*. Thus when a high diver jumps from a diving board, his moment of inertia, I , can be decreased by curling his body more, in which case his angular velocity ω is increased. Fig. 3.9 (ii). He may then be able to turn more somersaults before striking the water. Similarly, a dancer on skates can spin faster by folding her arms.

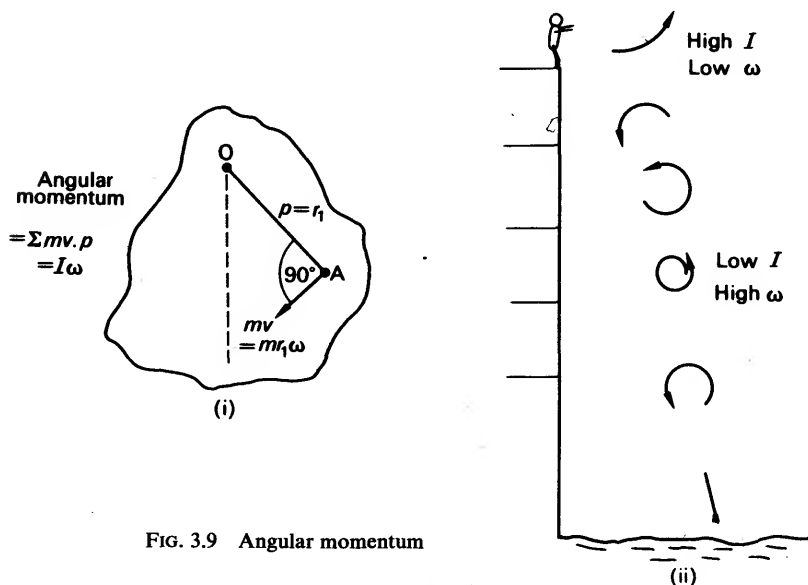


FIG. 3.9 Angular momentum

The earth is an object which rotates about an axis passing through its geographic north and south poles with a period of 1 day. If it is struck by meteorites, then, since action and reaction are equal, no external couple acts on the earth and meteorites. Their total angular momentum is thus conserved. Neglecting the angular momentum of the meteorites about the earth's axis before collision compared with that of the earth, then

$$\text{angular momentum of earth plus meteorites after collision} = \text{angular momentum of earth before collision.}$$

Since the effective mass of the earth has increased after collision the moment of inertia has increased. Hence the earth will slow up slightly.

Similarly, if a mass is dropped gently on to a turntable rotating freely at a steady speed, the conservation of angular momentum leads to a reduction in the speed of the table.

Angular momentum, and the principle of the conservation of angular momentum, have wide applications in physics. They are used in connection with enormous rotating masses such as the earth, as well as minute spinning particles such as electrons, neutrons and protons found inside atoms.

Experiment on Conservation of Angular Momentum

A simple experiment on the principle of the conservation of angular momentum is illustrated below.

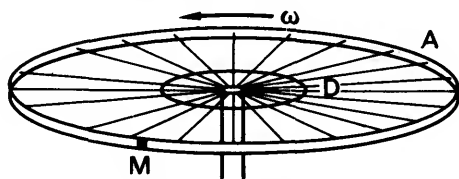


FIG. 3.10 Conservation of angular momentum

Briefly, in Fig. 3.10 (i) a bicycle wheel A without a tyre is set rotating in a horizontal plane and the time for three complete revolutions is obtained with the aid of a white tape marker M on the rim. A ring D of known moment of inertia, I , is then gently placed on the wheel concentric with it, by 'dropping' it from a small height. The time for the next three revolutions is then determined. This is repeated with several more rings of greater known moment of inertia.

If the principle of conservation of angular momentum is true, then $I_0\omega_0 = (I_0 + I_1)\omega_1$, where I_0 is the moment of inertia of the wheel alone, ω_0 is the angular frequency of the wheel alone, and ω_1 is the angular frequency with a ring. Thus t_0, t_1 are the respective times for three revolutions,

$$\frac{I_0 + I_1}{t_1} = \frac{I_0}{t_0}$$

$$\therefore \frac{I_1}{I_0} + 1 = \frac{t_1}{t_0}$$

Thus a graph of t_1/t_0 v. I_1 should be a straight line. Within the limits of experimental error, this is found to be the case.

EXAMPLE

Consider a disc of mass 100 g and radius 10 cm is rotating freely about axis O through its centre at 40 r.p.m. Fig. 3.11. Then, about O,

$$\text{moment of inertia } I = \frac{Ma^2}{2} = \frac{1}{2} \times 0.1 \text{ (kg)} \times 0.1^2 \text{ (m}^2\text{)} = 5 \times 10^{-4} \text{ kg m}^2,$$

and angular momentum = $I\omega = 5 \times 10^{-4}\omega$,
where ω is the angular velocity corresponding to 40 r.p.m.

Suppose some wax W of mass m 20 g is dropped gently on to the disc at a distance r of 8 cm from the centre O . The disc then slows down to another speed, corresponding to an angular velocity ω_1 say. The total angular momentum about O of disc plus wax

$$= I\omega_1 + mr^2\omega_1 = 5 \times 10^{-4}\omega_1 + 0.02 \times 0.08^2 \cdot \omega_1$$

$$= 6.28 \times 10^{-4}\omega_1.$$

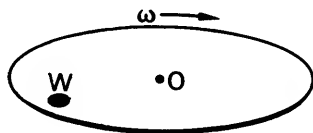


FIG. 3.11 Example

From the conservation of angular momentum for the disc and wax about O

$$6.28 \times 10^{-4}\omega_1 = 5 \times 10^{-4}\omega.$$

$$\therefore \frac{\omega_1}{\omega} = \frac{500}{628} = \frac{n}{40},$$

where n is the r.p.m. of the disc.

$$\therefore n = \frac{500}{628} \times 40 = 32 \text{ (approx).}$$

Kepler's law and angular momentum

Consider a planet moving in an orbit round the sun S . Fig. 3.12.

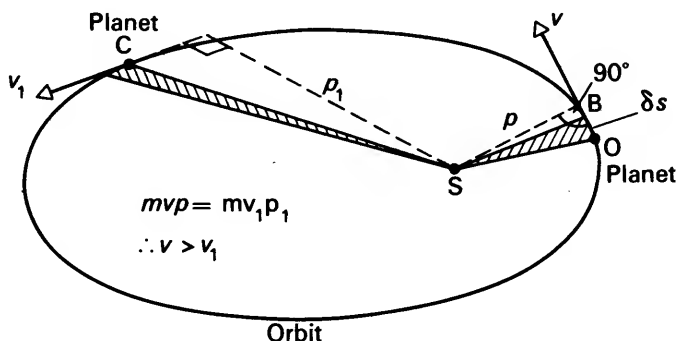


FIG. 3.12 Angular momentum and planets

At an instant when the planet is at O , its velocity v is along the tangent to the orbit at O . Suppose the planet moves a very small distance δs from O to B in a small time δt , so that the velocity $v = \delta s / \delta t$ and its direction is practically along OB . Then, if the conservation of angular momentum is obeyed,

$$mv \times p = \text{constant},$$

where m is the mass of the planet and p is the perpendicular from S to OB produced.

$$\therefore \frac{m \cdot \delta s \cdot p}{\delta t} = \text{constant}.$$

But the area δA of the triangle SBO = $\frac{1}{2}$ base \times height = $\delta s \times p/2$.

$$\therefore m \cdot 2 \frac{\delta A}{\delta t} = \text{constant}$$

$$\therefore \frac{\delta A}{\delta t} = \text{constant},$$

since $2m$ is constant. Thus if the conservation of angular momentum is true, the area swept out per second by the radius SO is constant while the planet O moves in its orbit. In other words, equal areas are swept out in equal times. *But this is Kepler's second law*, which has been observed to be true for centuries (see p. 58). Consequently, the principle of the conservation of angular momentum has stood the test of time. From the equality of the angular momentum values at O and C, where p is less than p_1 , it follows that v is greater than v_1 . Thus the planet speeds up on approaching S.

The force on O is always one of attraction towards S. It is described as a *central force*. Thus the force has no moment about O and hence the angular momentum of the planet about S is conserved.

Kinetic Energy of a Rolling Object

When an object such as a cylinder or ball rolls on a plane, the object is rotating as well as moving bodily along the plane; therefore it has rotational energy as well as translational energy.

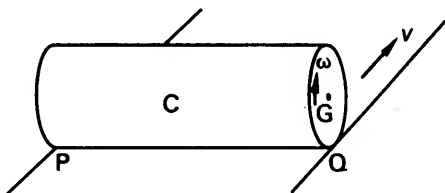


FIG. 3.13 Rolling object

Consider a cylinder C rolling along a plane without slipping, Fig. 3.13. At any instant the line of contact, PQ, with the plane is at rest, and we can consider the whole of the cylinder to be rotating about this axis. Hence the energy of the cylinder = $\frac{1}{2}I_1\omega^2$, where I_1 is the moment of inertia about PQ and ω is the angular velocity.

But if I is the moment of inertia about a parallel axis through the centre of gravity of the cylinder, M is the mass of the cylinder and a its radius, then

$$I_1 = I + Ma^2,$$

from the result on p. 79.

$$\begin{aligned}\therefore \text{energy of cylinder} &= \frac{1}{2}(I + Ma^2)\omega^2 \\ &= \frac{1}{2}I\omega^2 + \frac{1}{2}Ma^2\omega^2\end{aligned}$$

$$\therefore \text{Energy} = \frac{1}{2}I\omega^2 + \frac{1}{2}Mv^2 \quad \quad \quad (1)$$

since, by considering the distance rolled and the angle then turned, $v = a\omega =$ velocity of centre of gravity. This energy formula is true for any moving object.

As an application of the energy formula, suppose a *ring* rolls along a plane. The moment of inertia about the centre of gravity, its centre, $= Ma^2$ (p. 76); also, the angular velocity, ω , about its centre $= v/a$, where v is the velocity of the centre of gravity.

$$\begin{aligned}\therefore \text{kinetic energy of ring} &= \frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2 \\ &= \frac{1}{2}Mv^2 + \frac{1}{2}Ma^2 \times \left(\frac{v}{a}\right)^2 = Mv^2.\end{aligned}$$

By similar reasoning, the kinetic energy of a sphere rolling down a plane

$$\begin{aligned}&= \frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2 \\ &= \frac{1}{2}Mv^2 + \frac{1}{2} \times \frac{2}{5}Ma^2 \times \left(\frac{v}{a}\right)^2 = \frac{7}{10}Mv^2,\end{aligned}$$

since $I = 2Ma^2/5$ (p. 78).

Acceleration of Rolling Object

We can now deduce the acceleration of a rolling object down an inclined plane.

As an illustration, suppose a solid cylinder rolls down a plane. Then

$$\text{kinetic energy} = \frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2.$$

But moment of inertia, I , about an axis through the centre of gravity parallel to the plane $= \frac{1}{2}Ma^2$, and $\omega = v/a$, where a is the radius.

$$\therefore \text{kinetic energy} = \frac{1}{2}Mv^2 + \frac{1}{4}Mv^2 = \frac{3}{4}Mv^2.$$

If the cylinder rolls from *rest* through a distance s , the loss of potential energy $= Mgs \sin \alpha$, where α is the inclination of the plane to the horizontal.

$$\therefore \frac{3}{4}Mv^2 = Mgs \sin \alpha$$

$$\therefore v^2 = \frac{4g}{3}s \sin \alpha$$

But

$$v^2 = 2as, \text{ where } a \text{ is the linear acceleration.}$$

$$\therefore 2as = \frac{4g}{3}s \sin \alpha$$

$$\therefore a = \frac{2g}{3} \sin \alpha \quad \quad \quad (i)$$

The acceleration if sliding, and no rolling, took place down the plane is $g \sin \alpha$. The cylinder has thus a smaller acceleration when rolling.

The time t taken to move through a distance s from rest is given by $s = \frac{1}{2}at^2$. Thus, from (i),

$$\begin{aligned}s &= \frac{1}{3}gt^2 \sin \alpha, \\ \text{or } t &= \sqrt{\frac{3s}{g \sin \alpha}}.\end{aligned}$$

If the cylinder is *hollow*, instead of solid as assumed, the moment of inertia about an axis through the centre of gravity parallel to the plane is greater than that for a solid cylinder, assuming the same mass and dimensions (p. 88). The time taken for a hollow cylinder to roll a given distance from rest on the plane is then greater than that taken by the solid cylinder, from reasoning similar to that above; and thus if no other means were available, a time test on an inclined plane will distinguish between a solid and a hollow cylinder of the same dimensions and mass. If a torsion wire is available, however, the cylinders can be suspended in turn, and the period of torsional oscillations determined. The cylinder of larger moment of inertia, the hollow cylinder, will have a greater period, as explained on p. 89.

Measurement of Moment of Inertia of Flywheel

The moment of inertia of a flywheel W about a horizontal axle A can be determined by tying one end of some string to a pin on the axle, winding the string round the axle, and attaching a mass M to the other end of the string, Fig. 3.14. The length of string is such that M reaches the floor, when released, at the same instant as the string is completely unwound from the axle.

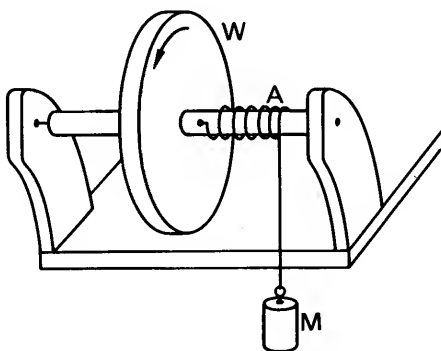


FIG. 3.14 Moment of inertia of flywheel

M is released, and the number of revolutions, n , made by the wheel W up to the occasion when M strikes the ground is noted. The further number of revolutions n_1 made by W until it comes finally to rest, and the time t taken, are also observed by means of a chalk-mark on W .

Now the loss in potential energy of M = gain in kinetic energy of M + gain in kinetic energy of flywheel + work done against friction.

$$\therefore Mgh = \frac{1}{2}Mr^2\omega^2 + \frac{1}{2}I\omega^2 + nf, \quad (i)$$

where h is the distance M has fallen, r is the radius of the axle, ω is the angular velocity, I is the moment of inertia, and f is the energy per turn expended against friction. Since the energy of rotation of the flywheel

when the mass M reaches the ground = work done against friction in n_1 revolutions, then

$$\frac{1}{2}I\omega^2 = n_1 f.$$

$$\therefore f = \frac{1}{2} \frac{I\omega^2}{n_1}.$$

Substituting for f in (i),

$$\therefore Mgh = \frac{1}{2}Mr^2\omega^2 + \frac{1}{2}I\omega^2 \left(1 + \frac{n}{n_1}\right) \quad (ii)$$

Since the angular velocity of the wheel when M reaches the ground is ω , and the final angular velocity of the wheel is zero after a time t , the average angular velocity = $\omega/2 = 2\pi n_1/t$. Thus $\omega = 4\pi n_1/t$. Knowing ω and the magnitude of the other quantities in (ii), the moment of inertia I of the fly-wheel can be calculated.

Period of Oscillation of Rigid Body

On p. 81 we showed that the moment of the forces acting on rotating objects = $Id\omega/dt = Id^2\theta/dt^2$, where I is the moment of inertia about the axis concerned and $d^2\theta/dt^2$ is the angular acceleration about the axis. Consider a rigid body oscillating about a fixed axis O , Fig. 3.15. The moment of the weight mg (the only external force) about O is $mgh \sin \theta$, or $mgh\theta$ if θ is small, where h is the distance of the centre of gravity from O .

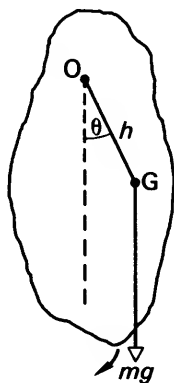


FIG. 3.15
Compound
pendulum

$$\therefore I \frac{d^2\theta}{dt^2} = -mgh\theta,$$

the minus indicating that the moment due to the weight always opposes the growth of the angle θ .

$$\therefore \frac{d^2\theta}{dt^2} = \frac{-mgh}{I} \theta = -\omega^2 \theta,$$

where $\omega^2 = mgh/I$.

\therefore the motion is simple harmonic motion (p. 44),

and period, $T = \frac{2\pi}{\omega} = \frac{2\pi}{\sqrt{mgh/I}} = 2\pi \sqrt{\frac{I}{mgh}} \quad (1)$

If $I = mk_1^2$, where k_1 is the radius of gyration about O ,

$$T = 2\pi \sqrt{\frac{mk_1^2}{mgh}} = 2\pi \sqrt{\frac{k_1^2}{gh}} \quad (2)$$

A ring of mass m and radius a will thus oscillate about an axis through a point O on its circumference normal to the plane of the ring with a period T given by

$$T = 2\pi \sqrt{\frac{I_0}{mga}}.$$

But $I_0 = I_G + ma^2$ (theorem of parallel axes) $= ma^2 + ma^2 = 2ma^2$.

$$\therefore T = 2\pi \sqrt{\frac{2ma^2}{mga}} = 2\pi \sqrt{\frac{2a}{g}}.$$

Thus if $a = 0.5$ m, $g = 9.8$ m s⁻²,

$$T = 2\pi \sqrt{\frac{2 \times 0.5}{9.8}} = 2.0 \text{ seconds (approx).}$$

Measurement of Moment of Inertia of Plate

The moment of inertia of a circular disc or other plate about an axis perpendicular to its plane, for example, can be measured by means of torsional oscillations. The plate is suspended horizontally from a vertical torsion wire, and the period T_1 of torsional oscillations is measured. Then, from (1),

$$T_1 = 2\pi \sqrt{\frac{I_1}{c}}, \quad \dots \dots \dots (i)$$

where I_1 is the moment of inertia and c is the constant (opposing couple per unit radian) of the wire (p. 164). A ring or annulus of *known* moment of inertia I_2 is now placed on the plate concentric with the axis, and the new period T_2 is observed. Then

$$T_2 = 2\pi \sqrt{\frac{I_1 + I_2}{c}} \quad \dots \dots \dots (ii)$$

By squaring (i) and (ii), and then eliminating c , we obtain

$$I_1 = \frac{T_1^2}{T_2^2 - T_1^2} \cdot I_2.$$

Thus knowing T_1 , T_2 , and I_2 , the moment of inertia I_1 can be calculated.

Compound Pendulum. Since $I = I_G + mh^2 = mk^2 + mh^2$, where I_G is the moment of inertia about the centre of gravity, h is the distance of the axis O from the centre of gravity, and k is the radius of gyration about the centre of gravity, then, from previous,

$$T = 2\pi \sqrt{\frac{I}{mgh}} = 2\pi \sqrt{\frac{mk^2 + mh^2}{mgh}}.$$

$$\therefore T = 2\pi \sqrt{\frac{k^2 + h^2}{gh}}.$$

Hence

$$T = 2\pi \sqrt{\frac{l}{g}}$$

where

$$l = \frac{k^2 + h^2}{h} \quad \dots \dots \dots (i)$$

Thus $(k^2 + h^2)/h$ is the length, l , of the equivalent simple pendulum.

From (i),

$$h^2 - hl + k^2 = 0.$$

$$\therefore h_1 + h_2 = l, \text{ and } h_1 h_2 = k^2,$$

where h_1 and h_2 are the roots of the equation.

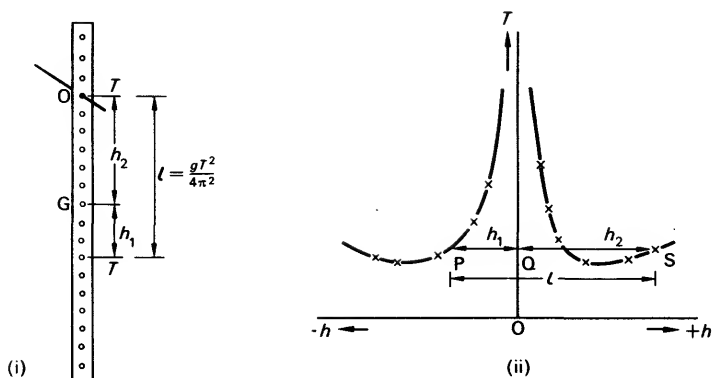


FIG. 3.16 Compound pendulum experiment

By timing the period of vibration, T , of a long rod about a series of axes at varying distances h on either side of the centre of gravity, and then plotting a graph of T v. h , two different values of h giving the same period can be obtained, Fig. 3.16 (i), (ii). Suppose h_1, h_2 are the two values. Then from the result just obtained, $h_1 + h_2 = l$, the length of the equivalent simple pendulum. Thus, since $T = 2\pi\sqrt{l/g}$,

$$g = \frac{4\pi^2 l}{T^2} = \frac{4\pi^2 (h_1 + h_2)}{T^2}.$$

In Fig. 3.16 (ii), $PQ + QS = h_1 + h_2 = l$.

Kater's Pendulum. The acceleration due to gravity was first measured by the simple pendulum method, and calculated from the relation $g = 4\pi^2 l/T^2$, with the usual notation. The length l , the distance from the point of suspension to the centre of gravity of the bob, however, cannot be determined with very great accuracy.

In 1817 Captain Kater designed a reversible pendulum, with knife-edges for the suspension; it was a compound pendulum. Now it has just been shown that the same period is obtained between two non-symmetrical points on a compound pendulum when their distance apart is l , the length of the equivalent simple pendulum. Thus if T is the period about either knife-edge when this occurs, $g = 4\pi^2 l/T^2$, where l is now the distance between the knife-edges. The pendulum is made geometrically symmetrical about the mid-point, with a brass bob at one end and a wooden bob of the same size at the other. A movable large and small weight are placed between the knife-edges, which are about one metre apart. The period is then slightly greater than 2 seconds.

To find g , the pendulum is set up in front of an accurate seconds clock, with the bob of the clock and that of the Kater pendulum in line with each other, and both sighted through a telescope. The large weight on the pendulum is moved until the period is nearly the same about either knife-edge, and the small weight is used as a fine adjustment. When the periods are the same, the distance l between the knife-edges is measured very accurately by a comparator method with a microscope and standard metre. Thus knowing T and l , g can be calculated from $g = 4\pi^2 l/T^2$.

For details of the experiment the reader should consult *Advanced Practical Physics for Students* by Worsnop and Flint (Methuen).

Summary

The following table compares the translational (linear) motion of a small mass m with the rotational motion of a large object of moment of inertia I .

Linear Motion	Rotational Motion
1. Velocity, v	Velocity $v = r\omega$
2. Momentum = mv	Angular momentum = $I\omega$
3. Energy = $\frac{1}{2}mv^2$	Rotational energy = $\frac{1}{2}I\omega^2$
4. Force, $F, = ma$	Torque, $C, = I \times \text{ang. accn. } (d^2\theta/dt^2)$
5. Simple pendulum: $T = 2\pi\sqrt{\frac{l}{g}}$	Compound pendulum: $T = 2\pi\sqrt{\frac{I}{mgh}}$
6. Motion down inclined plane—energy equation: $\frac{1}{2}mv^2 = mgh \sin \theta$	Rotating without slipping down inclined plane—energy equation: $\frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2 = Mgh \sin \theta$
7. Conservation of linear momentum on collision, if no external forces	Conservation of angular momentum on collision, if no external couple

EXAMPLE

What is meant by the *moment of inertia* of an object about an axis?

Describe and give the theory of an experiment to determine the moment of inertia of a flywheel mounted on a horizontal axle.

A uniform circular disc of mass 20 kg and radius 15 cm is mounted on a horizontal cylindrical axle of radius 1.5 cm and negligible mass. Neglecting frictional losses in the bearings, calculate (a) the angular velocity acquired from rest by the application for 12 seconds of a force of 2.0 kgf tangential to the axle, (b) the kinetic energy of the disc at the end of this period, (c) the time required to bring the disc to rest if a braking force of 0.1 kgf were applied tangentially to its rim. (L)

Moment of inertia of disc, $I, = \frac{1}{2}Ma^2 = \frac{1}{2} \times 20 \text{ (kg)} \times 0.15^2 \text{ (m}^2\text{)} = 0.225 \text{ kg m}^2$.

(a) Torque due to 2 kgf tangential to axle

$$= 2 \times 9.8 \text{ (N)} \times 0.015 \text{ (m)} = 0.294 \text{ N m.}$$

$$\therefore \text{angular acceleration} = \frac{\text{torque}}{I} = \frac{0.294}{0.225} \text{ rad s}^{-2}.$$

$$\therefore \text{after 12 seconds, angular velocity} = \frac{12 \times 0.294}{0.225} = 15.7 \text{ rad s}^{-1}.$$

(b) K.E. of disc after 12 seconds = $\frac{1}{2}I\omega^2$

$$= \frac{1}{2} \times 0.225 \times 15.7^2 = 27.8 \text{ J.}$$

(c) Decelerating torque = $0.1 \times 9.8 \text{ (N)} \times 0.15 \text{ (m)}$.

$$\therefore \text{angular deceleration} = \frac{\text{torque}}{I} = \frac{0.1 \times 9.8 \times 0.15}{0.225} \text{ rad s}^{-2}.$$

$$\therefore \text{time to bring disc to rest} = \frac{\text{initial angular velocity}}{\text{angular deceleration}}$$

$$= \frac{15.7 \times 0.225}{0.1 \times 9.8 \times 0.15} = 24 \text{ seconds.}$$

EXERCISES 3

(Assume $g = 10 \text{ m s}^{-2}$ unless otherwise stated)

What are the missing words in the statements 1–4?

1. The kinetic energy of an object rotating about an axis is calculated from ...
2. The angular momentum of the object is calculated from ...
3. ' $I \times$ angular acceleration' is equal to the ... on the object.
4. The period of oscillation of an object about an axis is calculated from ...

Which of the following answers, A, B, C, D or E, do you consider is the correct one in the statements 5–8?

5. When a sphere of moment of inertia I about its centre of gravity, and mass m , rolls from rest down an inclined plane without slipping, its kinetic energy is calculated from $A \frac{1}{2} I \omega^2$, $B \frac{1}{2} m v^2$, $C I \omega + m v$, $D \frac{1}{2} I \omega^2 + \frac{1}{2} m v^2$, $E I \omega$.

6. If a hoop of radius a oscillates about an axis through its circumference perpendicular to its plane, the period is $A 2\pi\sqrt{a/g}$, $B 2\pi\sqrt{2a/g}$, $C 2\pi\sqrt{g/a}$, $D 2\pi\sqrt{g/2a}$, $E a/2$.

7. Planets moving in orbit round the sun A increase in velocity at points near the sun because their angular momentum is constant, B increase in velocity near the sun because their energy is constant, C decrease in velocity near the sun owing to the increased attraction, D sweep out equal mass in equal times because their energy is constant, E always have circular orbits.

8. If a constant couple of 500 newton metre turns a wheel of moment of inertia 100 kg m^2 about an axis through its centre, the angular velocity gained in two seconds is $A 5 \text{ rad s}^{-1}$, $B 100 \text{ m s}^{-1}$, $C 200 \text{ m s}^{-1}$, $D 2 \text{ m s}^{-1}$, $E 10 \text{ rad s}^{-1}$.

9. A uniform rod has a mass of 60 g and a length 20 cm. Calculate the moment of inertia about an axis perpendicular to its length (i) through its centre, (ii) through one end. Prove the formulae used.

10. What is the *Theorem of Parallel Axes*? A uniform disc has a mass of 4 kg and a radius of 2 m. Calculate the moment of inertia about an axis perpendicular to its plane (i) through its centre, (ii) through a point of its circumference.

11. What is the *Theorem of Perpendicular Axes*? A ring has a radius of 20 cm and a mass of 100 g. Calculate the moment of inertia about an axis (i) perpendicular to its plane through its centre, (ii) perpendicular to its plane passing through a point on its circumference, (iii) in its plane passing through the centre.

12. What is the formula for the kinetic energy of (i) a particle, (ii) a rigid body rotating about an axis through its centre of gravity, (iii) a rigid body rotating about an axis through any point? Calculate the kinetic energy of a disc of mass 5 kg and radius 1 m rolling along a plane with a uniform velocity of 2 m s^{-1} .

13. A sphere rolls down a plane inclined at 30° to the horizontal. Find the acceleration and velocity of the sphere after it has moved 5.0 m from rest along the plane, assuming the moment of inertia of a sphere about a diameter is $\frac{2}{5} M a^2$, where M is the mass and a is the radius.

14. A uniform rod of length 3.0 m is suspended at one end so that it can move about an axis perpendicular to its length, and is held inclined at 60° to the vertical and then released. Calculate the angular velocity of the rod when (i) it is inclined at 30° to the vertical, (ii) it reaches the vertical.

15. Define the *moment of inertia* of a rigid object about an axis.

A ring of radius 2.0 m oscillates about an axis on its circumference which is perpendicular to the plane of the ring. Calculate the period of oscillation. Give an explanation of any formula used.

16. A flywheel with an axle 1.0 cm in diameter is mounted in frictionless bearings and set in motion by applying a steady tension of 200 gf to a thin thread wound tightly round the axle. The moment of inertia of the system about its axis of rotation is $5.0 \times 10^{-4} \text{ kg m}^2$. Calculate (a) the angular acceleration of the flywheel when 1 m of thread has been pulled off the axle, (b) the constant retarding couple which must then be applied to bring the flywheel to rest in one complete turn, the tension in the thread having been completely removed. (N.)

17. Define the moment of inertia of a body about a given axis. Describe how the moment of inertia of a flywheel can be determined experimentally.

A horizontal disc rotating freely about a vertical axis makes 100 r.p.m. A small piece of wax of mass 10 g falls vertically on to the disc and adheres to it at a distance of 9 cm from the axis. If the number of revolutions per minute is thereby reduced to 90, calculate the moment of inertia of the disc. (N.)

18. Describe an experiment using a bar pendulum to determine the acceleration due to gravity. Show how the result is calculated from the observations.

A uniform disc of diameter 12.0 cm and mass 810 g is suspended with its plane horizontal by a torsion wire and allowed to perform small torsional oscillations about a vertical axis through its centre. The disc is then replaced by a uniform sphere which is allowed to oscillate similarly about a diameter. If the period of oscillation of the sphere is 1.66 times that of the disc, determine the moment of inertia of the sphere about a diameter. (L.)

19. A plane sheet of metal of uniform thickness and of irregular shape is pierced with a number of small holes, irregularly distributed, so that it can be pivoted about an axis through any one of them to swing in its own plane. Describe how you would proceed in order to find the value of the moment of inertia of the sheet about an axis through its centre of gravity normal to its plane.

A rigid bar pendulum is pivoted at a distance h from its centre of gravity. When a piece of lead of negligible size and of mass M equal to that of the pendulum, is attached at the centre of gravity the periodic time of the pendulum is reduced to 0.80 of its former value. Find an expression for the moment of inertia of the pendulum about the pivot. (L.)

20. Define *moment of inertia* and derive an expression for the kinetic energy of a rigid body of moment of inertia I about a given axis when it is rotating about that axis with a uniform angular velocity ω .

Give two examples of physical phenomena in which moment of inertia is a necessary concept for a theoretical description, in each case showing how the concept is applied.

A uniform spherical ball starts from rest and rolls freely without slipping down an inclined plane at 10° to the horizontal along a line of greatest slope. Calculate its velocity after it has travelled 5 m. (M.I. of sphere about a diameter = $\frac{2Mr^2}{5}$.) (O. & C.)

21. Explain the meaning of the term *moment of inertia*. Describe in detail how you would find experimentally the moment of inertia of a bicycle wheel about the central line of its hub.

A uniform cylinder 20 cm long, suspended by a steel wire attached to its mid-point so that its long axis is horizontal, is found to oscillate with a period of

2 seconds when the wire is twisted and released. When a small thin disc, of mass 10 g, is attached to each end the period is found to be 2.3 seconds. Calculate the moment of inertia of the cylinder about the axis of oscillation. (N.)

22. What is meant by 'moment of inertia'? Explain the importance of this concept in dealing with problems concerning rotating bodies.

Describe, with practical details, how you would determine whether a given cylindrical body were hollow or not without damaging it. (C.)

23. Define *moment of inertia*, and find an expression for the kinetic energy of a rigid body rotating about a fixed axis.

A sphere, starting from rest, rolls (without slipping) down a rough plane inclined to the horizontal at an angle of 30° , and it is found to travel a distance of 13.5 m in the first 3 seconds of its motion. Assuming that F , the frictional resistance to the motion, is independent of the speed, calculate the ratio of F to the *weight* of the sphere. (For a sphere of mass m and radius r , the moment of inertia about a diameter is $\frac{2}{5}mr^2$.) (O. & C.)

chapter four

Static Bodies. Fluids

STATIC BODIES

Statics

1. STATICS is a subject which concerns the *equilibrium* of forces, such as the forces which act on a bridge. In Fig. 4.1 (i), for example, the joint O of a light bridge is in equilibrium under the action of the two forces P, Q acting in the girders meeting at O and the reaction S of the masonry at O.

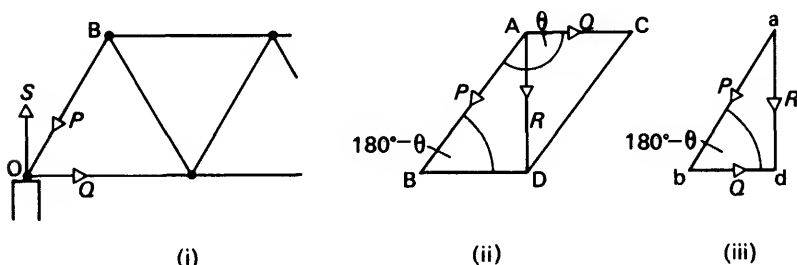


FIG. 4.1 Equilibrium of forces

Parallelogram of Forces

A force is a vector quantity, i.e., it can be represented in magnitude and direction by a straight line (p. 1). If AB, AC represent the forces P, Q respectively at the joint O, their *resultant*, R , is represented in magnitude and direction by the diagonal AD of the parallelogram $ABDC$ which has AB, AC as two of its adjacent sides, Fig. 4.1 (ii). This is known as the *parallelogram of forces*, and is exactly analogous to the parallelogram of velocities discussed on p. 8. Alternatively, a line ab may be drawn to represent the vector P and bd to represent Q , in which case ad represents the resultant R .

By trigonometry for triangle ABD , we have

$$AD^2 = BA^2 + BD^2 - 2BA \cdot BD \cos ABD.$$

$$\therefore R^2 = P^2 + Q^2 + 2PQ \cos \theta,$$

where θ = angle BAC ; the angle between the forces P, Q , = 180° - angle ABD . This formula enables R to be calculated when P, Q

and the angle between them are known. The angle BAD, or α say, between the resultant R and the force P can then be found from the relation

$$\frac{R}{\sin \theta} = \frac{Q}{\sin \alpha'}$$

applying the sine rule to triangle ABD and noting that angle ABD = $180^\circ - \theta$,

Resolved component. On p. 8 we saw that the effective part, or resolved component, of a vector quantity X in a direction θ inclined to it is given by $X \cos \theta$. Thus the resolved component of a force P in a direction making an angle of 30° with it is $P \cos 30^\circ$; in a perpendicular direction to the latter the resolved component is $P \cos 60^\circ$, or $P \sin 30^\circ$. In Fig. 4.1 (i), the downward component of the force P on the joint of O is given by $P \cos \text{BOS}$.

Forces in Equilibrium. Triangle of Forces

Since the joint O is in equilibrium, Fig. 4.1 (i), the resultant of the forces P , Q in the rods meeting at this joint is equal and opposite to the reaction S at O. Now the diagonal AD of the parallelogram ABDC in Fig. 4.1 (ii) represents the resultant R of P , Q since ABDC is the parallelogram of forces for P , Q ; and hence DA represents the force S . Consequently the sides of the triangle ABD represent the three forces at O in magnitude and direction:

This result can be generalised as follows. *If three forces are in equilibrium, they can be represented by the three sides of a triangle taken in order.* This theorem in Statics is known as the *triangle of forces*. In Fig. 4.1 (ii), AB, BD, DA, in this order, represent, P , Q , S respectively in Fig. 4.1 (i)

We can derive another relation between forces in equilibrium. Suppose X, Y are the respective algebraic sums of the resolved components in two perpendicular directions of three

forces P , Q , T in equilibrium, Fig. 4.2. Then, since X , Y can each be represented by the sides of a *rectangle* drawn to scale, their resultant R is given by

$$R^2 = X^2 + Y^2 \quad . \quad . \quad . \quad . \quad (i)$$

Now if the forces are in equilibrium, R is zero. It then follows from (i) that X must be zero and Y must be zero. Thus *if forces are in equilibrium the algebraic sum of their resolved components in any two perpendicular directions is respectively zero*. This result applies to any number of forces in equilibrium.

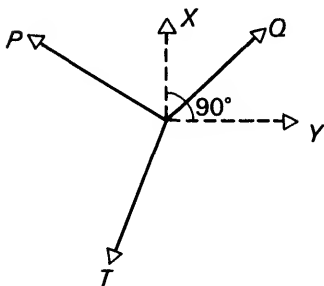


FIG. 4.2 Resolution of forces

EXAMPLE

State what is meant by *scalar* and *vector* quantities, giving examples of each.

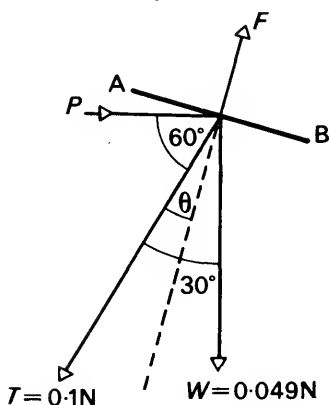


FIG. 4.3 Example

Explain how a flat kite can be flown in a wind that is blowing horizontally. The line makes an angle of 30° with the vertical and is under a tension of 0.1 newton; the mass of the kite is 5 g. What angle will the plane of the kite make with the vertical, and what force will the wind exert on it? (O. & C.)

Second part. When the kite AB is inclined to the horizontal, the wind blowing horizontally exerts an upward force F normal to AB, Fig. 4.3. For equilibrium of the kite, F must be equal and opposite to the resultant of the tension T , 0.1 newton, and the weight W , 0.005×9.8 or 0.049 N. By drawing the parallelogram of forces for the resultant of T and W , F and the angle θ between F and T can be found. θ is nearly 10° , and F is about 0.148 N. The angle between AB and the vertical = $60^\circ + \theta = 70^\circ$ (approx.). Also,

since F is the component of the horizontal force P of the wind.

$$P \cos (60^\circ + \theta) = F.$$

$$\therefore P = \frac{F}{\cos (60^\circ + \theta)} = \frac{0.148}{\cos 70^\circ} \\ = 0.43 \text{ N.}$$

Moments

When the steering-wheel of a car is turned, the applied force is said to exert a *moment*, or turning-effect, about the axle attached to the wheel. The magnitude of the moment of a force P about a point O is defined as *the product of the force P and the perpendicular distance OA* for all the forces in Fig. 4.4 (ii), we have

$$\text{moment} = P \times AO.$$

The magnitude of the moment is expressed in *newton metre* (N m) when

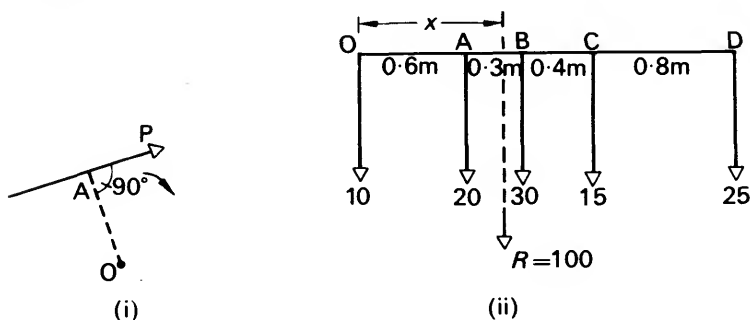


FIG. 4.4 Parallel forces

P is in newtons and AO is in metres. We shall take an anticlockwise moment as positive in sign and a clockwise moment as negative in sign.

Parallel Forces

If a rod carries loads of 10, 20, 30, 15, and 25 N at O, A, B, C, D respectively, the resultant R of the weights, which are parallel forces, for all the forces in Fig. 4.4 (ii), we have

$$\text{resultant, } R, = 10 + 20 + 30 + 15 + 25 = 100 \text{ N.}$$

Experiment and theory show that *the moment of the resultant of a number of forces about any point is equal to the algebraic sum of the moments of the individual forces about the same point*. This result enables us to find where the resultant R acts. Taking moments about O for all the forces in Fig. 4.4 (ii), we have

$$(20 \times 0.6) + (30 \times 0.9) + (15 \times 1.3) + (25 \times 2.1),$$

because the distances between the forces are 0.6 m, 0.3 m, 0.4 m, 0.8 m, as shown. If x m is the distance of the line of action of R from O, the moment of R about O = $R \times x = 100 \times x$.

$$\therefore 100x = (20 \times 0.6) + (30 \times 0.9) + (15 \times 1.3) + (25 \times 2.1),$$

from which

$$x = 1.1 \text{ m.}$$

Equilibrium of Parallel Forces

The resultant of a number of forces *in equilibrium* is zero; and the moment of the resultant about any point is hence zero. It therefore follows that the algebraic sum of the moments of all the forces about any point is zero when those forces are in equilibrium. This means that the total clockwise moment of the forces about any point = the total anticlockwise moment of the remaining forces about the same point.

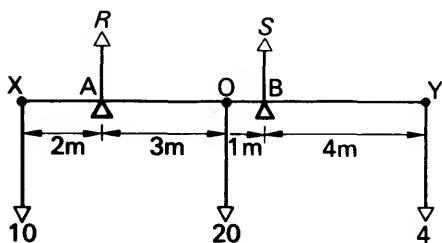


FIG. 4.5 Example

As a simple example of the equilibrium of parallel forces, suppose a light beam XY rests on supports, A, B, and has loads of 10, 20, and 4 N concentrated at X, O, Y respectively, Fig. 4.5. Let R , S be the reactions at A, B respectively. Then, for equilibrium in a vertical direction,

$$R + S = 10 + 20 + 4 = 34 \text{ N} \quad (i)$$

To find R , we take moments about a suitable point such as B , in which case the moment of S is zero. Then, for the remaining four forces,

$$+10 \cdot 6 + 20 \cdot 1 - R \cdot 4 - 4 \cdot 4 = 0,$$

from which $R = 16$ N. From (i), it follows that $S = 34 - 16 = 18$ N.

Equilibrium of Three Coplanar Forces

If any object is in equilibrium under the action of *three* forces, the resultant of two of the forces must be equal and opposite to the third force. Thus the line of action of the third force must pass through the point of intersection of the lines of action of the other two forces.

As an example of calculating unknown forces in this case, suppose that a 12 m ladder of 20.0 kgf is placed at an angle of 60° to the horizontal, with one end B leaning against a smooth wall and the other end

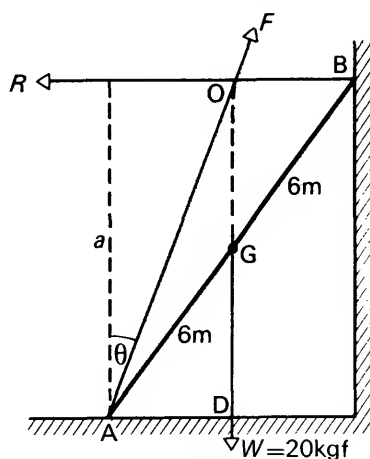


FIG. 4.6 Triangle of forces

A on the ground, Fig. 4.6. The force R at B on the ladder is called the *reaction* of the wall, and if the latter is smooth, R acts perpendicularly to the wall. Assuming the weight, W , of the ladder acts at its mid-point G , the forces W and R meet at O , as shown. Consequently the frictional force F at A passes through O .

The *triangle of forces* can be used to find the unknown forces R , F . Since DA is parallel to R , AO is parallel to F , and OD is parallel to W , the triangle of forces is represented by AOD . By means of a scale drawing R and F can be found, since

$$\frac{W(20)}{OD} = \frac{F}{AO} = \frac{R}{DA}.$$

A quicker method is to take moments about A for all the forces. The algebraic sum of the moments is zero about any point since the object is in equilibrium, and hence

$$R \cdot a - W \cdot AD = 0,$$

where a is the perpendicular from A to R . (F has zero moment about A .) But $a = 12 \sin 60^\circ$, and $AD = 6 \cos 60^\circ$.

$$\therefore R \times 12 \sin 60^\circ - 20 \times 6 \cos 60^\circ = 0.$$

$$\therefore R = 10 \frac{\cos 60^\circ}{\sin 60^\circ} = 5.8 \text{ kgf.}$$

Suppose θ is the angle F makes with the vertical.

Resolving the forces vertically, $F \cos \theta = W = 20 \text{ kgf.}$

Resolving horizontally, $F \sin \theta = R = 5.8 \text{ kgf.}$

$$\therefore F^2 \cos^2 \theta + F^2 \sin^2 \theta = F^2 = 20^2 + 5.8^2.$$

$$\therefore F = \sqrt{20^2 + 5.8^2} = 20.8 \text{ kgf.}$$

Couples and Torque

There are many examples in practice where two forces, acting together, exert a moment or turning-effect on some object. As a very simple case, suppose two strings are tied to a wheel at X , Y , and two equal and opposite forces, F , are exerted tangentially to the wheel, Fig. 4.7 (i). If the wheel is pivoted at its centre, O , it begins to rotate about O in an anticlockwise direction.

Two equal and opposite forces whose lines of action do not coincide are said to constitute a *couple* in Mechanics. The two forces always have a turning-effect, or moment, called a *torque*, which is defined by

$$\text{torque} = \text{one force} \times \text{perpendicular distance between forces} \quad (1)$$

Since XY is perpendicular to each of the forces F in Fig. 4.6 (i), the

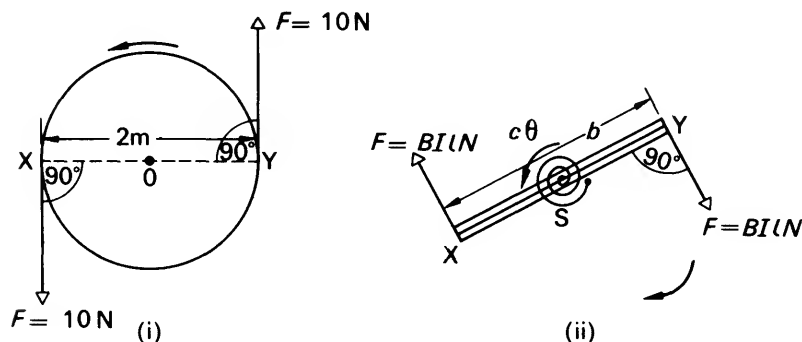


FIG. 4.7 Couple and torque

moment of the couple acting on the wheel $= F \times XY = F \times \text{diameter of wheel}$. Thus if $F = 10$ newton and the diameter is 2 metre, the moment of the couple or torque $= 20$ newton metre (N m).

In the theory of the *moving-coil electrical instrument*, we meet a case where a coil rotates when a current I is passed into it and comes to rest

after deflection through an angle θ . Fig. 4.7 (ii). The forces F on the two sides X and Y of the coil are both equal to $BIlN$, where B is the strength of the magnetic field, l is the length of the coil and N is the number of turns (see Electricity section, chapter 35). Thus the coil is deflected by a couple. The moment or torque of the deflecting couple $= F \times b$, where $b = XY =$ breadth of coil. Hence

$$\text{torque} = BI l N \times b = BAN I,$$

where $A = lb =$ area of coil. The opposing couple, due to the spring S, is $c\theta$, where c is its elastic constant (p.164). Thus, for equilibrium, $BANI = c\theta$.

Work Done by a Couple

Suppose two equal and opposite forces F act tangentially to a wheel W, and rotate it through an angle θ while the forces keep tangentially to the wheel, Fig. 4.8. The moment of the couple is then constant.

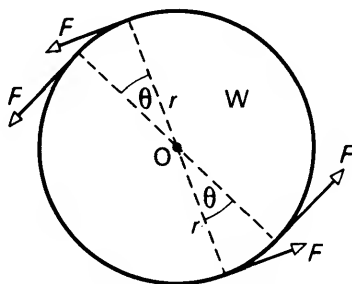


FIG. 4.8 Work done by couple

The work done by each force $= F \times \text{distance} = F \times r\theta$, since $r\theta$ is the distance moved by a point on the rim if θ is in radians.

$$\therefore \text{total work done by couple} = Fr\theta + Fr\theta = 2Fr\theta.$$

But $\text{moment of couple} = F \times 2r = 2Fr$

$$\therefore \text{work done by couple} = \text{torque or moment of couple} \times \theta$$

Although we have chosen a simple case, the result for the work done by a couple is always given by *torque \times angle of rotation*. In the formula, it should be carefully noted that θ is in radians. Thus suppose $F = 100 \text{ gf} = 0.1 \text{ kgf} = 0.1 \times 9.8 \text{ newton}$, $r = 4 \text{ cm} = 0.04 \text{ metre}$, and the wheel makes 5 revolutions while the moment of the couple is kept constant. Then

$$\text{torque or moment of couple} = 0.1 \times 9.8 \times 0.08 \text{ newton metre,}$$

$$\text{and angle of rotation} = 2\pi \times 5 \text{ radian.}$$

$$\therefore \text{work done} = 0.1 \times 9.8 \times 0.08 \times 2\pi \times 5 = 2.5 \text{ J}$$

Centre of Gravity

Every particle is attracted towards the centre of the earth by the force of gravity, and the *centre of gravity* of a body is the point where the *resultant* force of attraction or *weight* of the body acts. In the simple case of a ruler, the centre of gravity is the point of support when the ruler is balanced. A similar method can be used to find roughly the centre of gravity of a flat plate. A more accurate method consists of suspending the object in turn from two points on it, so that it hangs freely in each case, and finding the point of intersection of a plumb-line, suspended in turn from each point of suspension. This experiment is described in elementary books.

An object can be considered to consist of many small particles. The forces on the particles due to the attraction of the earth are all parallel since they act vertically, and hence their resultant is the sum of all the forces. The resultant is the *weight* of the whole object, of course. In the case of a rod of uniform cross-sectional area, the weight of a particle A at one end, and that of a corresponding particle A' at the other end, have a resultant which acts at the mid-point O of the rod, Fig. 4.9 (i).

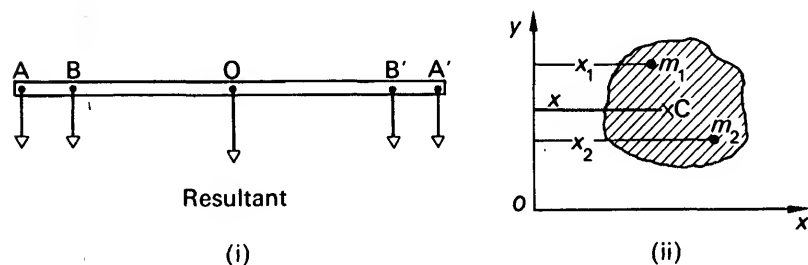


FIG. 4.9 Centre of gravity and mass

Similarly, the resultant of the weight of a particle B, and that of a corresponding particle at B', have a resultant acting at O. In this way, i.e., by symmetry, it follows that the resultant of the weights of all the particles of the rod acts at O. Hence the centre of gravity of a uniform rod is at its mid-point.

The centre of gravity, C.G., of the curved surface of a hollow cylinder acts at the midpoint of the cylinder axis. This is also the position of the C.G. of a uniform solid cylinder. The C.G. of a triangular plate or lamina is two-thirds of the distance along a median from corresponding point of the triangle. The C.G. of a uniform right solid cone is three-quarters along the axis from the apex.

Centre of Mass

The 'centre of mass' of an object is the point where its total mass acts or appears to act. Fig. 4.9 (ii) illustrates how the position of the centre of mass of an object may be calculated, using axes Ox , Oy .

If m_1 is the mass of a small part of the object and x_1 is the perpendicular distance to Oy , then m_1x_1 represents a product similar to the

moment of a weight at m_1 about Oy . Likewise, m_2x_2 is a 'moment' about Oy , where m_2 is another small part of the object. The sum of the total 'moments' about Oy of all the parts of the object can be written Σmx . The total mass $= \Sigma m = M$ say. The distance \bar{x} of the centre of mass C from Oy is then given by

$$\bar{x} = \frac{\Sigma mx}{M}.$$

Similarly, the distance \bar{y} of the centre of mass C from Ox is given by

$$\bar{y} = \frac{\Sigma my}{M}.$$

If the earth's field is uniform at all parts of the body, then the *weight* of a small mass m of it is typically mg . Thus, by moments, the distance of the centre of gravity from Oy is given by

$$\frac{\Sigma mg \times x}{\Sigma mg} = \frac{\Sigma mx}{\Sigma m} = \frac{\Sigma mx}{M}.$$

The acceleration due to gravity, g , cancels in numerator and denominator. It therefore follows that the centre of mass *coincides* with the centre of gravity. However, if the earth's field is *not* uniform at all parts of the object, the weight of a small mass m_1 is then m_1g_1 say and the weight of another small mass m_2 is m_2g_2 . Clearly, the centre of gravity does not now coincide with the centre of mass. A very long or large object has different values of g at various parts of it.

EXAMPLE

What is meant by (a) the centre of mass of a body, (b) the centre of gravity of a body?

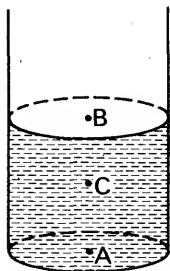


FIG. 4.10 Example

A cylindrical can is made of a material of mass 10 g cm^{-2} and has no lid. The diameter of the can is 25 cm and its height 50 cm . Find the position of the centre of mass when the can is half full of water. (C.)

The area of the base $= \pi r^2 = \pi \times (25/2)^2 \text{ cm}^2$; hence the mass is $\pi \times (25/2)^2 \times 10 \text{ g}$, and acts at A, the centre of the base, Fig. 4.10.

The mass of the curved surface of the can $= 2\pi rh \times 10 \text{ g} = 2\pi \times (25/2) \times 50 \times 10 \text{ g}$, and acts at B, half-way along the axis.

The mass of water $= \pi r^2 h \text{ g} = \pi \times (25/2)^2 \times 25 \text{ g}$, and acts at C, the mid-point of AB.

Thus the resultant mass in grammes

$$\begin{aligned} &= \frac{\pi \times 625 \times 10}{4} + \frac{2\pi \times 25 \times 50 \times 10}{2} + \frac{\pi \times 625 \times 25}{4} \\ &= \pi \times 625 \times 28\frac{3}{4}. \end{aligned}$$

Taking moments about A,

$$\therefore \pi \times 625 \times 28\frac{3}{4} \times x = (\pi \times 12500) \times AB + \left(\pi \times \frac{625 \times 25}{4} \right) \times AC$$

where x is the distance of the centre of mass from A.

$$\therefore 28\frac{3}{4}x = 20 \times 25 + \frac{25}{4} \times 12\frac{1}{2}$$

$$\therefore x = 20 \text{ (approx).}$$

\therefore centre of mass is 20 cm from the base.

Types of Equilibrium

If a marble A is placed on the curved surface of a bowl S, it rolls

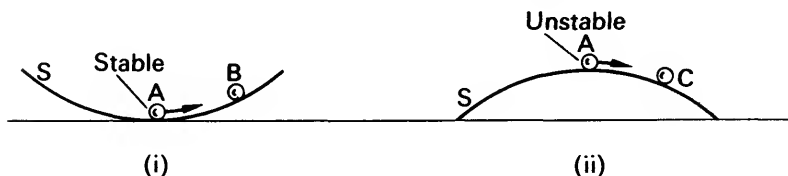


FIG. 4.11 Stable and unstable equilibrium

down and settles in equilibrium at the lowest point. Fig. 4.11 (i). Its potential energy is then a minimum. This is the case for objects in any field, gravitational, magnetic or electrical. The equilibrium position corresponds to minimum potential energy.

If the marble A is disturbed and displaced to B, its energy increases. When it is released, the marble rolls back to A. Thus the marble at A is said to be in *stable equilibrium*. Note that the centre of gravity of A is raised on displacement to B. On this account the forces in the field return the marble from B to A, where its potential energy is lower.

Suppose now that the bowl S is inverted and the marble is placed at its top point at A. Fig. 4.11 (ii). If A is displaced slightly to C, its potential energy and centre of gravity are then lowered. A now continues to move further away from B under the action of the forces in the field. Thus in Fig. 4.10 (ii), A is said to be in *unstable equilibrium*.

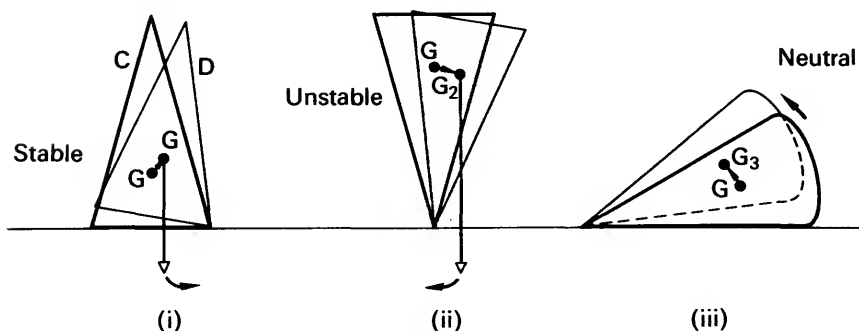


FIG. 4.12 Movement of C.G.

Fig. 4.12 (i) shows a cone C with its base on a horizontal surface. If it is slightly displaced to D, its centre of gravity G rises to G_1 . As previously explained, D returns to C when the cone is released, so that the equilibrium is stable. In Fig. 4.12 (ii), the cone is balanced on its apex. When it is slightly displaced, the centre of gravity, G_1 is lowered to G_2 . This is unstable equilibrium. Fig. 4.12 (iii) illustrates the case of the cone resting on its curved surface. If it is slightly displaced, the centre of gravity G remains at the same height G_3 . The cone hence remains in its displaced position. This is called *neutral equilibrium*.

EXAMPLE

A rectangular beam of thickness a is balanced on the curved surface of a rough cylinder of radius r . Show that the beam is stable if r is greater than $a/2$.

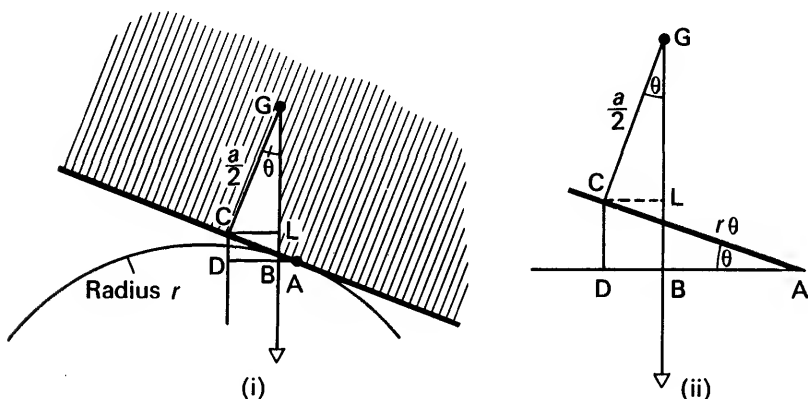


FIG. 4.13 Example

Suppose the beam is tilted through a small angle θ . The point of contact C then moves to A, the radius of the cylinder moves through an angle θ , and the vertical GB through the centre of gravity G of the beam makes an angle θ with CG. (Fig. 4.13 (i)). As shown in the exaggerated sketch in Fig. 4.13 (ii), $AC = r\theta$.

The beam is in stable equilibrium if the vertical through G lies to the left of A, since a restoring moment is then exerted. Thus for stable equilibrium, AD must be greater than DB, where CD is the vertical through C.

Now $AD = r\theta \cos \theta$, $DB = CL = \frac{a}{2} \sin \theta$.

$$\therefore r\theta \cos \theta > \frac{a}{2} \sin \theta.$$

When θ is very small, $\cos \theta \rightarrow 1$, $\sin \theta \rightarrow \theta$.

$$\therefore r\theta > \frac{a}{2} \theta.$$

$$\therefore r > \frac{a}{2}.$$

Common Balance

The common balance is basically a lever whose two arms are equal, Fig. 4.14. The fulcrum, about which the beam and pointer tilt, is an agate wedge resting on an agate plate; agate wedges, B, at the ends of the beam, support the scale-pans. The centre of gravity of the beam and pointer is vertically below the fulcrum, to make the arrangement stable. The weights placed on the two scale-pans are equal when there is a 'balance'.

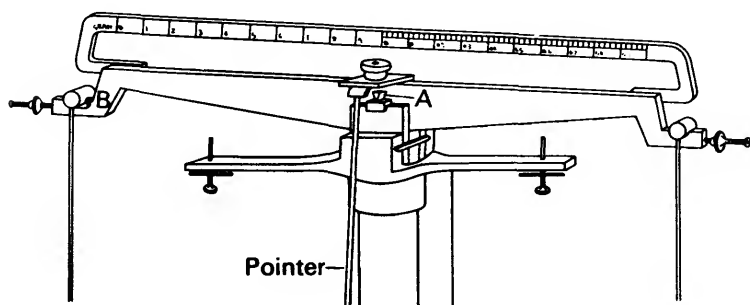


FIG. 4.14 Common balance

On rare occasions the arms of the balance are slightly unequal. The mass W of an object is then determined by finding the respective masses W_1 , W_2 required to balance it on each scale-pan. Suppose a , b are the lengths of the respective arms. Then, taking moments,

$$\therefore W_1 \cdot a = W \cdot b, \text{ and } W \cdot a = W_2 \cdot b.$$

$$\therefore \frac{W}{W_1} = \frac{a}{b} = \frac{W_2}{W}$$

$$\therefore W^2 = W_1 W_2$$

$$\therefore W = \sqrt{W_1 W_2}.$$

Thus W can be found from the two masses W_1 , W_2 .

Sensitivity of a Balance

A balance is said to be very *sensitive* if a small difference in weights on the scale-pans causes a large deflection of the beam. To investigate the factors which affect the sensitivity of a balance, suppose a weight W_1 is placed on the left scale-pan and a slightly smaller weight W_2 is placed on the right scale-pan, Fig. 4.15. The beam AOB will then be inclined at some angle θ to the horizontal, where O is the fulcrum.

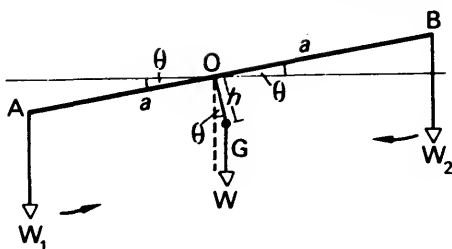


FIG. 4.15 Theory of balance

The weight W of the beam and pointer acts at G , at a distance h below O . Suppose $AO = OB = a$. Then, taking moments about O ,

$$W_1 a \cos \theta = Wh \sin \theta + W_2 a \cos \theta$$

$$\therefore (W_1 - W_2) a \cos \theta = Wh \sin \theta$$

$$\therefore \tan \theta = \frac{(W_1 - W_2)a}{Wh}$$

Thus for a given value of $(W_1 - W_2)$, the difference of the weights on the scale-pans, θ will increase when a increases and W, h both decrease. In theory, then, a sensitive balance must be light and have long arms, and the centre of gravity of its beam and pointer must be very close to the fulcrum. Now a light beam will not be rigid. Further, a beam with long arms will take a long time to settle down when it is deflected. A compromise must therefore be made between the requirements of sensitivity and those of design.

If the knife-edges of the scale-pan and beam are in the same plane, corresponding to A, B and O in Fig. 4.15, then the weights W_1, W_2 on them always have the same perpendicular distance from O , irrespective of the inclination of the beam. In this case the net moment about O is $(W_1 - W_2)a \cos \theta$. Thus the moment depends on the difference, $W_1 - W_2$, of the weights and not on their actual values. Hence the sensitivity is independent of the actual load value over a considerable range.

When the knife-edge of the beam is below the knife-edges of the two scale-pans, the sensitivity increases with the load; the reverse is the case if the knife-edge of the beam is above those of the scale-pans.

Buoyancy Correction in Weighing

In very accurate weighing, a correction must be made for the buoyancy of the air. Suppose the body weighed has a density ρ and a mass m . From Archimedes principle (p. 114), the upthrust due to the air of density σ is equal to the weight of air displaced by the body, and

hence the net downward force = $\left(m - \frac{m}{\rho} \cdot \sigma\right)g$, since the volume of the

body is m/ρ . Similarly, if the weights restoring a balance have a total mass m_1 and a density ρ_1 , the net downward force $= \left(m_1 - \frac{m_1}{\rho_1} \cdot \sigma\right)g$. Since there is a balance,

$$m - \frac{m\sigma}{\rho} = m_1 - \frac{m_1\sigma}{\rho_1}$$

$$\therefore m = m_1 \frac{\left(1 - \frac{\sigma}{\rho_1}\right)}{1 - \frac{\sigma}{\rho}}$$

Thus knowing the density of air, σ , and the densities ρ, ρ_1 , the true mass m can be found in terms of m_1 . The pressure and temperature of air, which may vary from day to day, affects the magnitude of its density σ , from the gas laws; the humidity of the air is also taken into account in very accurate weighing, as the density of moist air differs from that of dry air.

FLUIDS

Pressure

Liquids and gases are called *fluids*. Unlike solid objects, fluids can flow.

If a piece of cork is pushed below the surface of a pool of water and then released, the cork rises to the surface again. The liquid thus exerts an upward force on the cork and this is due to the *pressure* exerted on the cork by the surrounding liquid. Gases also exert pressures. For example, when a thin closed metal can is evacuated, it

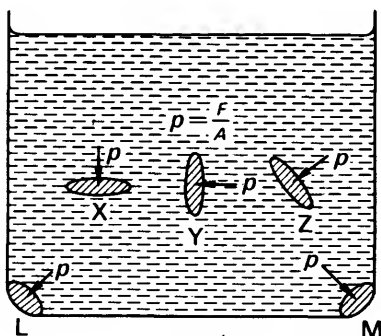


FIG. 4.16 Pressure in liquid

usually collapses with a loud explosion. The surrounding air now exerts a pressure on the outside which is no longer counter-balanced by the pressure inside, and hence there is a resultant force.

Pressure is defined as the *average force per unit area* at the particular region of liquid or gas. In Fig. 4.16, for example, X represents a small

horizontal area, Y a small vertical area and Z a small inclined area, all inside a vessel containing a liquid. The pressure p acts normally to the planes of X, Y or Z. In each case

$$\text{average pressure, } p, = \frac{F}{A},$$

where F is the normal force due to the liquid on an area A of X, Y or Z. Similarly, the pressure p on the sides L or M of the curved vessel act normally to L and M have magnitude F/A . In the limit, when the area is very small, $p = dF/dA$.

At a given point in a liquid, the pressure can act in any direction. Thus *pressure is a scalar*, not a vector. The direction of the force on a particular surface is normal to the surface.

Formula for Pressure

Observation shows that the pressure increases with the depth, h , below the liquid surface and with its density ρ .

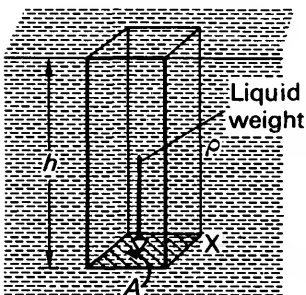


FIG. 4.17 Pressure and depth

To obtain a formula for the pressure, p , suppose that a horizontal plate X of area A is placed at a depth h below the liquid surface, Fig. 4.17. By drawing vertical lines from points on the perimeter of X, we can see that the force on X due to the liquid is equal to the weight of liquid of height h and uniform cross-section A . Since the volume of this liquid is Ah , the mass of the liquid $= Ah \times \rho$.

$$\therefore \text{weight} = Ah\rho g \text{ newton,}$$

where g is 9.8, h is in m, A is in m^2 , and ρ is in kg m^{-3} .

$$\therefore \text{pressure, } p, \text{ on } X = \frac{\text{force}}{\text{area}} = \frac{Ah\rho g}{A}$$

$$\therefore p = h\rho g \quad (1)$$

When h , ρ , g have the units already mentioned, the pressure p is in newton m^{-2} (N m^{-2}).

1 bar = 10^6 dyne cm^{-2} . To change 10^6 dyne cm^{-2} to N m^{-2} , we may proceed as follows:

$$10^6 \frac{\text{dyne}}{\text{cm}^2} = 10^6 \frac{\text{dyne}}{\text{N}} \cdot \frac{\text{N}}{\text{m}^2} \cdot \frac{\text{m}^2}{\text{cm}^2} = 10^6 \times \frac{1}{10^5} \cdot \frac{\text{N}}{\text{m}^2} \cdot 10^4 = 10^5 \frac{\text{N}}{\text{m}^2}$$

$$\therefore 1 \text{ bar} = 10^5 \text{ N m}^{-2} \quad (2)$$

Pressure is often expressed in terms of that due to a height of mercury (Hg). One unit is the *torr* (after Torricelli):

$$1 \text{ torr} = 1 \text{ mmHg} = 133.3 \text{ N m}^{-2} \text{ (approx).}$$

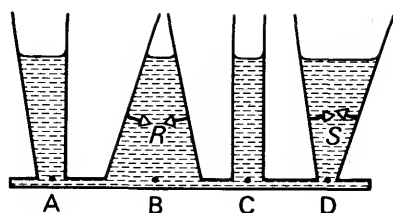


FIG. 4.18 Pressure and cross-section

From $p = h\rho g$ it follows that *the pressure in a liquid is the same at all points on the same horizontal level in it*. Experiment also gives the same result. Thus a liquid filling the vessel shown in Fig. 4.18 rises to the same height in each section if ABCD is horizontal. The cross-sectional area of B is greater than that of D; but the force on B is the sum of the weight of water above it together with the downward

component of reaction R of the sides of the vessel, whereas the force on D is the weight of water above it *minus* the upward component of the reaction S of the sides of the vessel. It will thus be noted that the pressure in a vessel is independent of the cross-sectional area of the vessel.

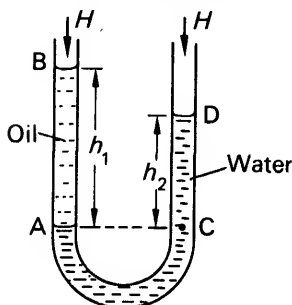


FIG. 4.19

Comparison of densities

Liquids in U-tube

Suppose a U-tube is partly filled with water, and oil is then poured into the left side of the tube. The oil will then reach some level B at a height h_1 above the surface of separation, A, of the water and oil, while the water on the right side of the tube will then reach some level D at a height h_2 above the level of A, Fig. 4.19.

Since the pressure in the water at A is equal to the pressure at C on the same horizontal level, it follows that

$$H + h_1\rho_1g = H + h_2\rho_2g,$$

where H is the atmospheric pressure, and ρ_1, ρ_2 are the respective densities of oil and water. Simplifying,

$$h_1\rho_1 = h_2\rho_2$$

$$\therefore \rho_1 = \rho_2 \times \frac{h_2}{h_1}.$$

Since ρ_2 (water) = 1000 kg m^{-3} , and h_2, h_1 can be measured, the density ρ_1 of the oil can be found.

Atmospheric Pressure

The pressure of the atmosphere was first measured by Galileo, who observed the height of a water column in a tube placed in a deep well.

About 1640 TORRICELLI thought of the idea of using mercury instead of water, to obtain a much shorter column. He completely filled a glass tube about a metre long with mercury, and then inverted it in a vessel D containing the liquid, taking care that no air entered the tube. He observed that the mercury in the tube fell to a level A about 76 cm or 0.76 m above the level of the mercury in D, Fig. 4.20. Since there was

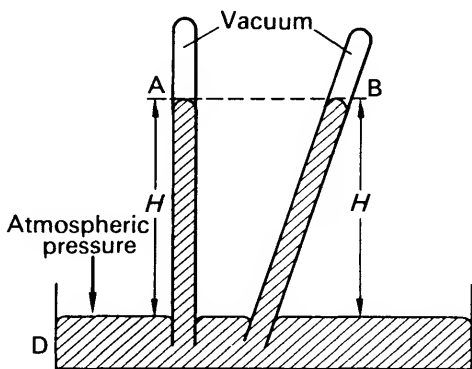


FIG. 4.20 Atmospheric pressure

no air originally in the tube, there must be a vacuum above the mercury at A, and it is called a *Torricellian vacuum*. This was the first occasion in the history of science that a vacuum had been created.

If the tube in Fig. 4.20 is inclined to the vertical, the mercury ascends the tube to a level B at the same vertical height H above the level of the mercury in D as A.

The pressure on the surface of the mercury in D is atmospheric pressure; and since the pressure is transmitted through the liquid, the atmospheric pressure supports the column of mercury in the tube. Suppose A is at a height H above the level of the mercury in D. Now the pressure, p , at the bottom of a column of liquid of height H and density ρ is given by $p = H\rho g$ (p. 110). Thus if $H = 760 \text{ mm} = 0.76 \text{ m}$ and $\rho = 13600 \text{ kg m}^{-3}$,

$$p = H\rho g = 0.76 \times 13600 \times 9.8 = 1.013 \times 10^5 \text{ newton metre}^{-2}.$$

The pressure at the bottom of a column of mercury 76 cm high for a particular mercury density and value of g is known as *standard pressure* or *one atmosphere*. By definition, 1 atmosphere = $1.01325 \times 10^5 \text{ N m}^{-2}$. *Standard temperature and pressure* (S.T.P.) is 0°C and 76 cm Hg pressure.

A *bar* is the name given to a pressure of one million (10^6) dyne cm^{-2} , and is thus very nearly equal to one atmosphere. 1 bar = 10^5 newton m^{-2} (p. 110).

Fortin's Barometer

A *barometer* is an instrument for measuring the pressure of the atmosphere, which is required in weather-forecasting, for example. The most accurate form of barometer is due to FORTIN, and like the simple arrangement already described, it consists basically of a barometer tube containing mercury, with a vacuum at the top, Fig. 4.21. One end of the tube dips into a pool of mercury contained in a washleather bag B. A brass scale C graduated in centimetres and millimetres is fixed at the top of the barometer. The zero of the scale correspondings to the tip of an ivory tooth P, and hence, before the level of the top of the mercury is read from the scales, the screw S is adjusted until the level of the mercury in B just reaches the tip of P. A vernier scale V can be moved by a screw D until the bottom of it just reaches the top of the mercury in the tube, and the reading of the height of the mercury is taken from C and V. Torricelli was the first person to observe the variation of the barometric height as the weather changed.

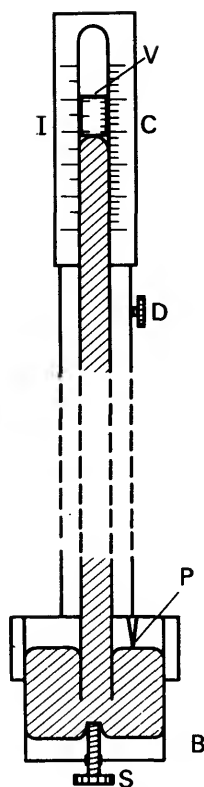


FIG. 4.21
Fortin barometer

'Correction' to the Barometric Height

For comparison purposes, the pressure read on a barometer is often 'reduced' or 'corrected' to the magnitude the pressure would have at 0°C and at sea-level, latitude 45° . Suppose the 'reduced' pressure is H_o cm of mercury, and the observed pressure is H_t cm of mercury, corresponding to a temperature of $t^{\circ}\text{C}$. Then, since pressure = $h\rho g$ (p. 110),

$$H_o\rho_o g = H_t\rho_t g',$$

where g is the acceleration due to gravity at sea-level, latitude 45° , and g' is the acceleration at the latitude of the place where the barometer was read.

$$\therefore H_o = H_t \times \frac{\rho_t}{\rho_o} \times \frac{g}{g'}$$

The magnitude of g'/g can be obtained from standard tables. The ratio ρ_t/ρ_o of the densities = $1/(1 + \gamma t)$, where γ is the absolute or true cubic expansivity of mercury. Further, the observed height H_t , on the brass scale requires correction for the expansion of brass from the temperature at which it was correctly calibrated. If the latter is 0°C , then the corrected height is $H_t(1 + \alpha t)$, where α is the mean linear

expansivity of brass. Thus, finally, the 'corrected' height H is given by

$$H_o = H_t \cdot \frac{1 + \alpha t}{1 + \gamma t} \cdot \frac{g'}{g}.$$

For further accuracy, a correction must be made for the surface tension of mercury (p. 132).

Variation of atmospheric pressure with height

The density of a liquid varies very slightly with pressure. The density of a gas, however, varies appreciably with pressure. Thus at sea-level the density of the atmosphere is about 1.2 kg m^{-3} ; at 1000 m above sea-level the density is about 1.1 kg m^{-3} ; and at 5000 m above sea-level it is about 0.7 kg m^{-3} . Normal atmospheric pressure is the pressure at the base of a column of mercury 760 mm high, a liquid which has a density of about 13600 kg m^{-3} . Suppose air has a constant density of about 1.2 kg m^{-3} . Then the height of an air column of this density which has a pressure equal to normal atmospheric pressure

$$= \frac{760}{1000} \times \frac{13600}{1.2} \text{ m} = 8.4 \text{ km}.$$

In fact, the air 'thins' the higher one goes, as explained above. The height of the air is thus much greater than 8.4 km.

Density, Relative Density

As we have seen, the pressure in a fluid depends on the density of the fluid.

The *density* of a substance is defined as its *mass per unit volume*. Thus

$$\text{density, } \rho, = \frac{\text{mass of substance}}{\text{volume of substance}} \quad (47)$$

The density of copper is about 9.0 g cm^{-3} or $9 \times 10^3 \text{ kg m}^{-3}$; the density of aluminium is 2.7 g cm^{-3} or $2.7 \times 10^3 \text{ kg m}^{-3}$; the density of water at 4°C is 1 g cm^{-3} or 1000 kg m^{-3} .

Substances which float on water have a density less than 1000 kg m^{-3} (p. 117). For example, ice has a density of about 900 kg m^{-3} ; cork has a density of about 250 kg m^{-3} . Steel, of density 8500 kg m^{-3} , will float on mercury, whose density is about 13600 kg m^{-3} at 0°C .

The density of a substance is often expressed relative to the density of water. This is called the *relative density* or *specific gravity* of the substance. It is a ratio or number, and has no units. The relative density of mercury is 13.6. Thus the density of mercury is 13.6 times the density of water, 1000 kg m^{-3} , and is hence 13600 kg m^{-3} . Copper has a relative density of 9.0 and hence a density of 9000 kg m^{-3} .

Archimedes' Principle

An object immersed in a fluid experiences a resultant upward force owing to the pressure of fluid on it. This upward force is called the *upthrust* of the fluid on the object. ARCHIMEDES stated that *the upthrust is equal to the weight of fluid displaced by the object*, and this is known as

Archimedes' Principle. Thus if an iron cube of volume 400 cm^3 is totally immersed in water of density 1 g cm^{-3} , the upthrust on the cube $= 400 \times 1 = 400 \text{ gf}$. If the same cube is totally immersed in oil of density 0.8 g cm^{-3} , the upthrust on it $= 400 \times 0.8 = 320 \text{ gf}$.

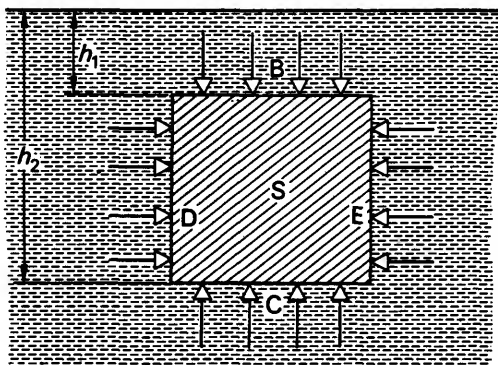


FIG. 4.22. Archimedes' Principle

Fig. 4.22 shows why Archimedes' Principle is true. If S is a solid immersed in a liquid, the pressure on the lower surface C is greater than on the upper surface B , since the pressure at the greater depth h_2 is more than that at h_1 . The pressure on the remaining surfaces D and E act as shown. The force on each of the four surfaces is calculated by summing the values of *pressure \times area* over every part, remembering that vector addition is needed to sum forces. With a simple *rectangular-shaped solid* and the sides, D , E vertical, it can be seen that (i) the resultant horizontal force is zero, (ii) the upward force on $C = \text{pressure} \times \text{area } A = h_2 \rho g A$, where ρ is the liquid density and the downward force on $B = \text{pressure} \times \text{area } A = h_1 \rho g A$. Thus

$$\text{resultant force on solid} = \text{upward force (upthrust)} = (h_2 - h_1) \rho g A.$$

But $(h_2 - h_1)A = \text{volume of solid, } V$,
 $\therefore \text{upthrust} = V \rho g = mg$, where $m = V \rho$.
 $\therefore \text{upthrust} = \text{weight of liquid displaced}.$

With a solid of irregular shape, taking into account horizontal and vertical components of forces, the same result is obtained. The upthrust is the weight of *liquid displaced* whatever the nature of the object immersed, or whether it is hollow or not. This is due primarily to the fact that the pressure on the object depends on the liquid in which it is placed.

Density or Relative Density measurement by Archimedes' Principle

The upthrust on an object immersed in water, for example, is the difference between (i) its weight in air when attached to a spring-balance and (ii) the reduced reading on the spring-balance or 'weight'

when it is totally immersed in the liquid. Suppose the upthrust is found to be 100 gf. Then, from Archimedes' Principle, the object displaces 100 gf of water. But the density of water is 1 g cm^{-3} . Hence the volume of the object = 100 cm^3 , which is numerically equal to the difference in weightings in (i) and (ii).

The density or relative density of a *solid* such as brass or iron can thus be determined by (1) weighing it in air, m_0 gf say, (2) weighing it when it is totally immersed in water, m_1 gf say. Then

$$\text{upthrust} = m_0 - m_1 = \text{wt. of water displaced.}$$

$$\therefore \text{relative density of solid} = \frac{m_0}{m_0 - m_1},$$

$$\text{and density of solid, } \rho_s = \frac{m_0}{m_0 - m_1} \times \text{density of water.}$$

The density or relative density of a *liquid* can be found by weighing a solid in air (m_0), then weighing it totally immersed in the liquid (m_1), and finally weighing it totally immersed in water (m_2).

Now $m_0 - m_2 = \text{upthrust in water} = \text{weight of water displaced,}$

and $m_0 - m_1 = \text{upthrust in liquid} = \text{weight of liquid displaced.}$

$$\therefore \frac{m_0 - m_1}{m_0 - m_2} = \text{relative density of liquid,}$$

$$\text{or } \frac{m_0 - m_1}{m_0 - m_2} \times \text{density of water} = \text{density of liquid.}$$

Density of Copper Sulphate crystals

If a solid dissolves in water, such as a copper sulphate crystal for example, its density can be found by totally immersing it in a liquid in which it is insoluble. Copper sulphate can be weighed in paraffin oil, for example. Suppose the apparent weight is m_1 , and the weight in air is m_0 . Then

$$m_0 - m_1 = \text{upthrust in liquid} = V\rho,$$

where V is the volume of the solid and ρ is the density of the liquid.

$$\therefore V = \frac{m_0 - m_1}{\rho}.$$

$$\therefore \text{density of solid} = \frac{\text{mass}}{\text{volume}} = \frac{m_0}{V} = \frac{m_0}{m_0 - m_1} \cdot \rho.$$

The density, ρ , of the liquid can be found by means of a density bottle, for example. Thus knowing m_0 and m_1 , the density of the solid can be calculated.

Density of Cork

If a solid floats in water, cork for example, its density can be found by attaching a brass weight or 'sinker' to it so that both solids become totally immersed in water. The apparent weight (m_1) of the sinker and cork together is then obtained. Suppose m_2 is the weight of the sinker in air, m_3 is the weight of the sinker alone in water, and m_0 is the weight of the cork in air.

Then

$$m_2 - m_3 = \text{upthrust on sinker in water.}$$

$$\therefore m_0 + m_2 - m_1 - (m_2 - m_3) = \text{upthrust on cork in water}$$

$$= m_0 - m_1 + m_3$$

$$\therefore \text{relative density of cork} = \frac{m_0}{m_0 - m_1 + m_3}.$$

Flotation

When an object *floats* in a liquid, the upthrust on the object must be equal to its weight for equilibrium. Cork has a density of about 0.25 g cm^{-3} , so that 100 cm^3 of cork has a mass of 25 g . In water, then, cork sinks until the upthrust is 25 gf . Now from Archimedes' Principle, 25 gf is the weight of water displaced. Thus the cork sinks until 25 cm^3 of its 100 cm^3 volume is immersed. The fraction of the volume immersed is hence equal to the relative density.

Ice has a density of about 0.9 g cm^{-3} . A block of ice therefore floats in water with about $\frac{9}{10}$ ths of it immersed.

Hydrometer

Hydrometers use the principle of flotation to measure density or relative density. Fig. 4.23. Since they have a constant weight, the upthrust when they float in a liquid is always the same. Thus in a liquid of density 1.0 g cm^{-3} , a hydrometer of 20 gf will sink until 20 cm^3 is immersed. In a liquid of 2.0 g cm^{-3} , it will sink until only 10 cm^3 is immersed. The density or relative density readings hence increase in a *downward* direction, as shown in Fig. 4.23.

Practical hydrometers have a weighted end M for stability, a wide bulb to produce sufficient upthrust to counterbalance the weight, and a narrow stem BL for sensitivity. If V is the whole volume of the hydrometer in Fig. 4.23, a is the area of the stem and y is the length not immersed in a liquid of density ρ , then

$$\text{upthrust} = \text{wt. of liquid displaced} = (V - ay)\rho = w,$$

where w is the weight of the hydrometer.

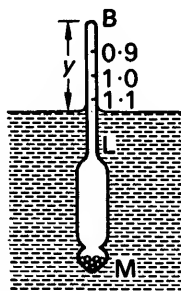


FIG. 4.23
Hydrometer

EXAMPLE

An ice cube of mass 50.0 g floats on the surface of a strong brine solution of volume 200.0 cm^3 inside a measuring cylinder. Calculate the level of the liquid in the measuring cylinder (i) before and (ii) after all the ice is melted. (iii) What happens to the level if the brine is replaced by 200.0 cm^3 water and 50.0 g of ice is again added? (Assume density of ice, brine = $900, 1100 \text{ kg m}^{-3}$ or $0.9, 1.1 \text{ g cm}^{-3}$.)

(i) Floating ice displaces 50 g of brine since upthrust is 50 gf .

$$\therefore \text{volume displaced} = \frac{\text{mass}}{\text{density}} = \frac{50}{1.1} = 45.5 \text{ cm}^3.$$

$$\therefore \text{level on measuring cylinder} = 245.5 \text{ cm}^3.$$

(ii) 50 g of ice forms 50 g of water when all of it is melted.

\therefore level on measuring cylinder falls to 250.0 cm^3 .

(iii) *Water.* Initially, volume of water displaced = 50 cm^3 , since upthrust = 50 g.

\therefore level on cylinder = 250.0 cm^3 .

If 1 g of ice melts, volume displaced is 1 cm^3 less. But volume of water formed is 1 cm^3 . Thus the net change in water level is zero. Hence the water level remains unchanged as the ice melts.

Fluids in Motion. Streamlines and velocity

A stream or river flows slowly when it runs through open country and faster through narrow openings or constrictions. As shown shortly, this is due to the fact that water is practically an incompressible fluid, that is, changes of pressure cause practically no change in fluid density at various parts.

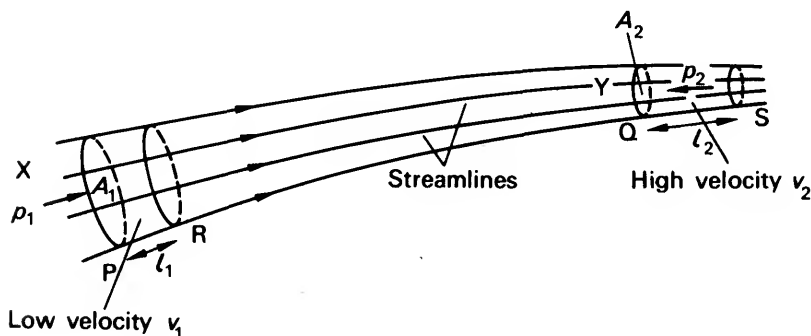


FIG. 4.24 Bernoulli's theorem

Fig. 4.24 shows a tube of water flowing steadily between X and Y, where X has a bigger cross-sectional area A_1 than the part Y, of cross-sectional area A_2 . The *streamlines* of the flow represent the directions of the velocities of the particles of the fluid and the flow is uniform or laminar (p. 204). Assuming the liquid is incompressible, then, if it moves from PQ to RS, the volume of liquid between P and R is equal to the volume between Q and S. Thus $A_1 l_1 = A_2 l_2$, where l_1 is PR and l_2 is QS, or $l_2/l_1 = A_1/A_2$. Hence l_2 is greater than l_1 . Consequently the *velocity* of the liquid at the narrow part of the tube, where, it should be noted, the streamlines are closer together, is greater than at the wider part Y, where the streamlines are further apart. For the same reason, slow-running water from a tap can be made into a fast jet by placing a finger over the tap to narrow the exit.

Pressure and velocity. Bernoulli's Principle

About 1740, Bernoulli obtained a relation between the pressure and velocity at different parts of a moving incompressible fluid. If the viscosity is negligibly small, there are no frictional forces to overcome (p. 174). In this case the work done by the pressure difference per unit

volume of a fluid flowing along a pipe steadily is equal to the gain of kinetic energy per unit volume plus the gain in potential energy per unit volume.

Now the work done by a pressure in moving a fluid through a distance = force \times distance moved = (pressure \times area) \times distance moved = pressure \times volume moved, assuming the area is constant at a particular place for a short time of flow. At the beginning of the pipe where the pressure is p_1 , the work done per unit volume on the fluid is thus p_1 ; at the other end, the work done per unit volume by the fluid is likewise p_2 . Hence the net work done on the fluid per unit volume = $p_1 - p_2$. The kinetic energy per unit volume = $\frac{1}{2}$ mass per unit volume \times velocity². = $\frac{1}{2}\rho \times \text{velocity}^2$, where ρ is the density of the fluid. Thus if v_2 and v_1 are the final and initial velocities respectively at the end and the beginning of the pipe, the kinetic energy gained per unit volume = $\frac{1}{2}\rho(v_2 - v_1)^2$. Further, if h_2 and h_1 are the respective heights measured from a fixed level at the end and beginning of the pipe, the potential energy gained per unit volume = mass per unit volume $\times g \times (h_2 - h_1)$ = $\rho g(h_2 - h_1)$.

Thus, from the conservation of energy,

$$\begin{aligned} p_1 - p_2 &= \frac{1}{2}\rho(v_2^2 - v_1^2) + \rho g(h_2 - h_1) \\ \therefore p_1 + \frac{1}{2}\rho v_1^2 + \rho g h_1 &= p_2 + \frac{1}{2}\rho v_2^2 + \rho g h_2 \\ \therefore p + \frac{1}{2}\rho v^2 + \rho g h &= \text{constant,} \end{aligned}$$

where p is the pressure at any part and v is the velocity there. Hence it can be said that, for streamline motion of an incompressible non-viscous fluid,

the sum of the pressure at any part plus the kinetic energy per unit volume plus the potential energy per unit volume there is always constant.

This is known as *Bernoulli's principle*.

Bernoulli's principle shows that at points in a moving fluid where the potential energy change $\rho g h$ is very small, or zero as in flow through a horizontal pipe, the pressure is low where the velocity is high; conversely, the pressure is high where the velocity is low. The principle has wide applications.

EXAMPLE

As a numerical illustration of the previous analysis, suppose the area of cross-section A_1 of X in Fig. 4.25 is 4 cm^2 , the area A_2 of Y is 1 cm^2 , and water flows past each section in laminar flow at the rate of $400 \text{ cm}^3 \text{ s}^{-1}$. Then

$$\text{at X, speed } v_1 \text{ of water} = \frac{\text{vol. per second}}{\text{area}} = 100 \text{ cm s}^{-1} = 1 \text{ m s}^{-1};$$

$$\text{at Y, speed } v_2 \text{ of water} = 400 \text{ cm s}^{-1} = 4 \text{ m s}^{-1}.$$

The density of water, $\rho = 1000 \text{ kg m}^{-3}$.

$$\therefore p = \frac{1}{2}\rho(v_2^2 - v_1^2) = \frac{1}{2} \times 1000 \times (4^2 - 1^2) = 7.5 \times 10^3 \text{ newton m}^{-2}.$$

If h is in metres, $\rho = 1000 \text{ kg m}^{-3}$ for water, $g = 9.8 \text{ m s}^{-2}$, then, from $h\rho g$,

$$h = \frac{7.5 \times 10^3}{1000 \times 9.8} = 0.77 \text{ m (approx.)}$$

The pressure head h is thus equivalent to 0.77 m of water.

Applications of Bernoulli's Principle

1. A suction effect is experienced by a person standing close to the platform at a station when a fast train passes. The fast-moving air between the person and train produces a decrease in pressure and the excess air pressure on the other side pushes the person towards the train.

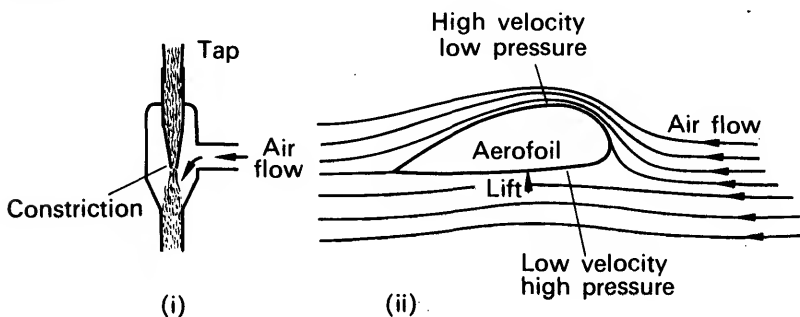


FIG. 4.25 Fluid velocity and pressure

2. *Filter pump.* A filter pump has a narrow section in the middle, so that a jet of water from the tap flows faster here. Fig. 4.25 (i). This causes a drop in pressure near it and air therefore flows in from the side tube to which a vessel is connected. The air and water together are expelled through the bottom of the filter pump.

3. *Aerofoil lift.* The curved shape of an aerofoil creates a fast flow of air over its top surface than the lower one. Fig. 4.25 (ii). This is shown by the closeness of the streamlines above the aerofoil compared with those below. From Bernoulli's principle, the pressure of the air below is greater than that above, and this produces the lift on the aerofoil.

4. *Flow of liquid from wide tank.* Suppose a liquid flows through a hole H at the bottom of a wide tank, as shown in Fig. 4.26. Assuming negligible viscosity and streamline flow at a small distance from the hole, which is an approximation, Bernoulli's theorem can be applied. At the top X of the liquid in the tank, the pressure is atmospheric, say B , the height measured from a fixed level such as the hole H is h , and the kinetic energy is negligible if the tank is wide so that the level falls very slowly. At the bottom, Y , near H , the pressure is again B , the height above H is now zero, and the kinetic energy is $\frac{1}{2}\rho v^2$, where ρ is the density and v is the

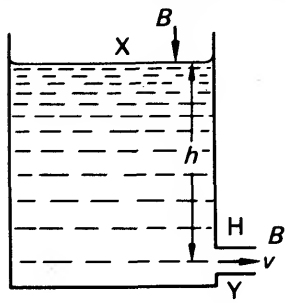


FIG. 4.26 Torricelli's theorem

velocity of emergence of the liquid. Thus, from Bernoulli's Principle,

$$B + \rho hg = B + \frac{1}{2} \rho v^2$$

$$\therefore v^2 = 2gh$$

Thus the velocity of the emerging liquid is the same as that which would be obtained if it fell freely through a height h , and this is known as *Torricelli's theorem*. In practice the velocity is less than that given by $\sqrt{2gh}$ owing to viscous forces, and the lack of streamline flow must also be taken into account.

EXERCISES 4

What are the missing words in the statements 1–6?

1. In SI units, the moment or torque of a couple is measured in ...
2. In stable equilibrium, when an object is slightly displaced its centre of gravity ...
3. When an object is in equilibrium under the action of three non-parallel forces, the three forces must ... one point.
4. The component of a force F in a direction inclined to it at an angle θ is ...
5. The sensitivity of a beam balance depends on the depth of the ... below the fulcrum.
6. When an object floats, the weight of fluid displaced is equal to the ...
7. In laminar flow of non-viscous fluid along a pipe, at regions of high pressure the ... is low.

Which of the following answers, A, B, C, D or E, do you consider is the correct one in the statements 8–10?

8. If a cone is balanced on its apex on a horizontal table and then slightly displaced, the potential energy of the cone is then A increased, B decreased, C constant, D a minimum, E a maximum.

9. If a hydrometer of mass 20 g and volume 30 cm³ has a graduated stem of 1 cm², and floats in water, the exposed length of stem is A 30 cm, B 25 cm, C 20 cm, D 10 cm, E 1 cm.

10. In laminar flow of a non-viscous fluid along a horizontal pipe, the work per second done by the pressure at any section is equal to A the pressure, B the volume per second there, C pressure \times volume per second there, D pressure \times volume, E pressure \times area of cross-section.

11. A flat plate is cut in the shape of a square of side 20.0 cm, with an equilateral triangle of side 20.0 cm adjacent to the square. Calculate the distance of the centre of mass from the apex of the triangle.

12. The foot of a uniform ladder is on a rough horizontal ground, and the top rests against a smooth vertical wall. The weight of the ladder is 40 kgf, and a man weighing 80 kgf stands on the ladder one-quarter of its length from the bottom. If the inclination of the ladder to the horizontal is 30°, find the reaction at the wall and the total force at the ground.

13. A rectangular plate ABCD has two forces of 10 kgf acting along AB and DC in opposite directions. If $AB = 3$ m, $BC = 5$ m, what is the moment of the couple acting on the plate? What forces acting along BC and AD respectively are required to keep the plate in equilibrium?

14. A hollow metal cylinder 2 m tall has a base of diameter 35 cm and is filled with water to a height of (i) 1 m, (ii) 50 cm. Calculate the distance of the centre of gravity in metre from the base in each case if the cylinder has no top. (Metal weighs 20 kg m^{-2} of surface. Assume $\pi = 22/7$.)

15. A trap-door 120 cm by 120 cm is kept horizontal by a string attached to the mid-point of the side opposite to that containing the hinge. The other end of the string is tied to a point 90 cm vertically above the hinge. If the trap-door weight is 5 kgf, calculate the tension in the string and the reaction at the hinge.

16. Two smooth inclined planes are arranged with their lower edges in contact; the angles of inclination of the plane to the horizontal are 30° , 60° respectively, and the surfaces of the planes are perpendicular to each other. If a uniform rod rests in the principal section of the planes with one end on each plane, find the angle of inclination of the rod to the horizontal.

17. Describe and give the theory of an accurate beam balance. Point out the factors which influence the sensitivity of the balance. Why is it necessary, in very accurate weighing, to take into account the pressure, temperature, and humidity of the atmosphere? (O. & C.)

18. Summarise the various conditions which are being satisfied when a body remains in equilibrium under the action of three non-parallel forces.

A wireless aerial attached to the top of a mast 20 m high exerts a horizontal force upon it of 60 kgf. The mast is supported by a stay-wire running to the ground from a point 6 m below the top of the mast, and inclined at 60° to the horizontal. Assuming that the action of the ground on the mast can be regarded as a single force, draw a diagram of the forces acting on the mast, and determine by measurement or by calculation the force in the stay-wire. (C.)

19. The beam of a balance has mass 150 g and its moment of inertia is $5 \times 10^{-4} \text{ kg m}^2$. Each arm of the balance is 10 cm long. When set swinging the beam makes one complete oscillation in 6 seconds. How far is the centre of gravity of the beam below its point of support, and through what angle would the beam be deflected by a weight of 1 milligram placed in one of the scale pans? (C.)

20. Under what conditions is a body said to be in equilibrium? What is meant by (a) *stable equilibrium* and (b) *unstable equilibrium*? Give one example of each.

A pair of railway carriage wheels, each of radius r , are joined by a thin axle; the mass of the whole is m . A light arm of length l ($l < r$) is attached perpendicularly to the axle and the free end of the arm carries a point mass M . The wheels rest, with the axle horizontal, on rails which are laid down a slope inclined at an angle ϕ to the horizontal. Show that, provided that ϕ is not too large and that the wheels do not slip on the rails, there are two values of the angle θ that the arm makes with the horizontal when the system is in equilibrium, and find these values of θ . Discuss whether, in each case, the equilibrium is stable or unstable. (O. & C.)

21. Give a labelled diagram to show the structure of a beam balance. Show if the knife-edges are collinear the sensitivity is independent of the load. Discuss other factors which then determine the sensitivity.

A body is weighed at a place on the equator, both with a beam balance and a very sensitive spring balance, with identical results. If the observations are

repeated at a place near one of the poles, using the same two instruments, discuss whether identical results will again be obtained. (L.)

22. Three forces in one plane act on a rigid body. What are the conditions for equilibrium?

The plane of a kite of mass 6 kg is inclined to the horizon at 60° . The resultant thrust of the air on the kite acts at a point 25 cm above its centre of gravity, and the string is attached at a point 30 cm above the centre of gravity. Find the thrust of the air on the kite, and the tension in the string. (C.)

23. In what circumstances is a physical system in equilibrium? Distinguish between stable, unstable and neutral equilibria.

Discuss the stability of the equilibrium of a uniform rough plank of thickness t , balanced horizontally on a rough cylindrical-fixed log of radius r , it being assumed that the axes of plank and log lie in perpendicular directions. (N.)

24. State the conditions of equilibrium for a body subjected to a system of coplanar parallel forces and briefly describe an experiment which you could carry out to verify these conditions.

Show how the equilibrium of a beam balance is achieved and discuss the factors which determine its sensitivity. Explain how the sensitivity of a given balance may be altered and why, for a particular adjustment, the sensitivity may be practically independent of the mass in the balance pans. Why is it inconvenient in practice to attempt to increase the sensitivity of a given balance beyond a certain limit? (O. & C.)

Fluids

25. An alloy of mass 588 g and volume 100 cm^3 is made of iron of relative density 8.0 and aluminium of relative density 2.7. Calculate the proportion (i) by volume, (ii) by mass of the constituents of the alloy.

26. A string supports a solid iron object of mass 180 g totally immersed in a liquid of density 800 kg m^{-3} . Calculate the tension in the string if the density of iron is 8000 kg m^{-3} .

27. A hydrometer floats in water with 6.0 cm of its graduated stem unimmersed, and in oil of relative density 0.8 with 4.0 cm of the stem unimmersed. What is the length of stem unimmersed when the hydrometer is placed in a liquid of relative density 0.9?

28. An alloy of mass 170 g has an apparent weight of 95 gf in a liquid of density 1.5 g cm^{-3} . If the two constituents of the alloy have relative densities of 4.0 and 3.0 respectively, calculate the proportion by volume of the constituents in the alloy.

29. State the principle of Archimedes and use it to derive an expression for the resultant force experienced by a body of weight W and density σ when it is totally immersed in a fluid of density ρ .

A solid weighs 237.5 g in air and 12.5 g when totally immersed in a liquid of relative density 0.9. Calculate (a) the specific gravity of the solid, (b) the relative density of a liquid in which the solid would float with one-fifth of its volume exposed above the liquid surface. (L.)

30. Distinguish between *mass* and *weight*. Define *density*.

Describe and explain how you would proceed to find an accurate value for the density of gold, the specimen available being a wedding ring of pure gold.

What will be the reading of (a) a mercury barometer, (b) a water barometer, when the atmospheric pressure is 10^5 N m^{-2} ? The density of mercury may be taken as 13600 kg m^{-3} and the pressure of saturated water vapour at room temperature as 13 mm of mercury. (L.)

31. Describe an experiment which demonstrates the difference between laminar and turbulent flow in a fluid.

A straight pipe of uniform radius R is joined, in the same straight line, to a narrower pipe of uniform radius r . Water (which may be assumed to be incompressible) flows from the wider into the narrower pipe. The velocity of flow in the wider pipe is V and in the narrower pipe is v . By equating work done against fluid pressures with change of kinetic energy of the water, show that the hydrostatic pressure is lower where the velocity of flow is higher.

Describe and explain one practical consequence or application of this difference in pressures. (O. & C.)

32. Describe some form of barometer used for the accurate measurement of atmospheric pressure, and point out the corrections to be applied to the observation.

Obtain an expression for the correction to be applied to the reading of a mercurial barometer when the reading is made at a temperature other than 0°C . (L.)

33. State the principle of Archimedes, and discuss its application to the determination of specific gravities by means of a common hydrometer. Why is this method essentially less accurate than the specific gravity bottle?

A common hydrometer is graduated to read specific gravities from 0.8 to 1.0. In order to extend its range a small weight is attached to the stem, above the liquid, so that the instrument reads 0.8 when floating in water. What will be the specific gravity of the liquid corresponding to the graduation 1.0? (O. & C.)

34. A hydrometer consists of a bulb of volume V and a uniform stem of volume v per cm of its length. It floats upright in water so that the bulb is just completely immersed. Explain for what density range this hydrometer may be used and how you would determine the density of such liquids. Describe the graph which would be obtained by plotting the reciprocal of the density against the length of the stem immersed.

A hydrometer such as that described sinks to the mark 3 on the stem, which is graduated in cm, when it is placed in a liquid of density 0.95 g cm^{-3} . If the volume per cm of the stem is 0.1 cm^3 , find the volume of the bulb. (L.)

35. A straight rod of length l , small cross sectional area a and of material density ρ is supported by a thread attached to its upper end. Initially the rod hangs in a vertical position over a liquid of density σ and then is lowered until it is partially submerged. Derive and discuss the equilibrium conditions of the rod neglecting surface tension. (N.)

chapter five

Surface Tension

Intermolecular Forces

THE forces which exist between molecules can explain many of the bulk properties of solids, liquids and gases. These intermolecular forces arise from two main causes:

(1) The *potential energy* of the molecules, which is due to interactions with surrounding molecules (this is principally electrical, not gravitational, in origin).

(2) The *thermal energy* of the molecules—this is the kinetic energy of the molecules and depends on the temperature of the substance concerned.

We shall see later that the particular state or phase in which matter appears—that is, solid, liquid or gas—and the properties it then has, are determined by the relative magnitudes of these two energies.

Potential energy and Force

In bulk, matter consists of numerous molecules. To simplify the situation, Fig. 5.1 shows the variation of the potential energy V between two molecules at a distance r apart.

Along the part BCD of the curve, the potential energy V is negative. Along the part AB, the potential energy V is positive. The force between the molecules is always given by $F = -dV/dr = -\text{potential gradient}$. Along CD the force is *attractive* and it decreases with distance r according to an inverse-power of r . Along ABC, the force is *repulsive*. Fig. 5.1 shows the variation of F with r .

At C, the minimum potential energy point of the curve, the molecules would be at their normal distance apart in the absence of thermal energy. The equilibrium distance OM, r_0 , is of the order 2 or 3×10^{-10} m (2 or 3 \AA) for a solid. At this distance apart, the attractive and repulsive forces balance each other. If the molecules are closer, ($r < r_0$), they would repel each other. If they are further apart, ($r > r_0$), they attract each other.

Phases or States of Matter

The molecules in a *solid* are said to be in a 'condensed' phase or state. Their thermal energy is then relatively low compared with their potential energy V and the molecules are 'bound' to each other. They may now vibrate about C, the minimum of the curve in Fig. 5.1.

When the thermal energy increases by an amount corresponding to CC' in Fig. 5.1, the molecule can then oscillate between the limits

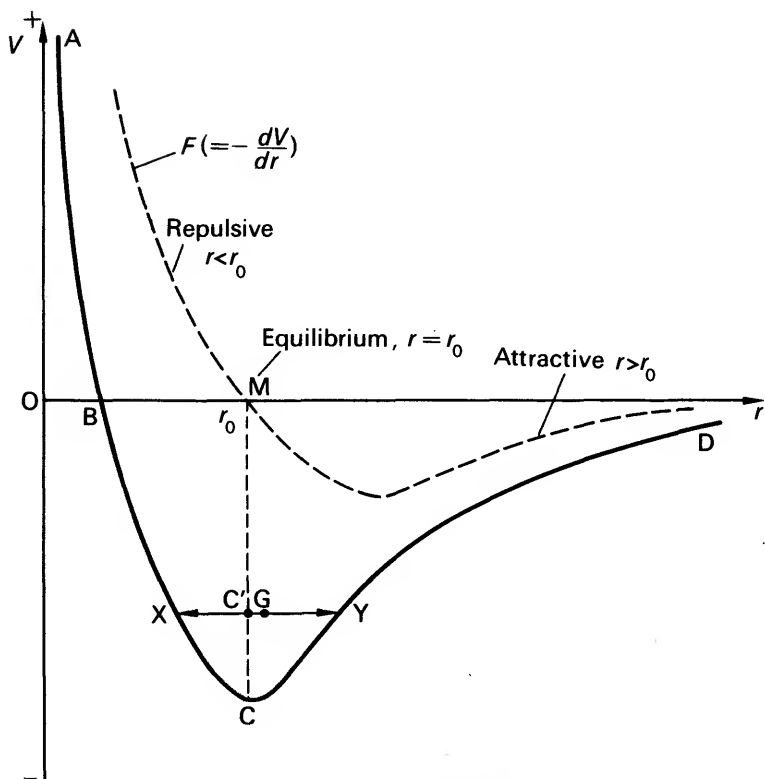


FIG. 5.1 Molecular potential energy and force

corresponding to X and Y . From the graph of force, F , it can be seen that when the molecule is on the left of C' it experiences a greater force towards it than when on the right. Consequently the molecule returns quicker to C' . Thus the *mean position* G is on the right of C' . This corresponds to a mean separation of molecules which is *greater* than r_0 . Thus the solid expands when its thermal energy is increased.

As the thermal energy increases further, at some particular temperature the molecules are able to move comparatively freely relative to neighbouring molecules. The solid then loses its rigid form and becomes a *liquid*. The molecules in the liquid constantly exchange places with other molecules, whereas in a solid the neighbours of a particular molecule remain unchanged. Further, the molecules of a liquid have translational as well as vibrational energy, that is, they move about constantly through the liquid, whereas molecules of a solid have vibrational energy only.

As the temperature of the liquid rises, the thermal energy of the molecules further increases. The average distance between the molecules then also increases and so their mean potential energy approaches zero, as can be seen from Fig. 5.1. At some stage the increased thermal

energy enables the molecules to completely break the bonds of attraction which keep them in a liquid state. The molecules then have little or no interaction and now form a *gas*. At normal pressures the forces of attraction between the gas molecules are comparatively very small and the molecules move about freely inside the volume they occupy. Gas molecules which are monatomic such as helium have translational energy only. Gas molecules such as oxygen or carbon dioxide, with two or more atoms, have rotational and vibrational energies in addition to translational energy.

Gases

At normal pressure, permanent gases such as air or oxygen obey Boyle's law, $pV = \text{constant}$, to a very good approximation. Now in the absence of attractive forces between the molecules, and assuming their actual volume is negligibly small, the kinetic theory of gases shows that Boyle's law is obeyed by this ideal gas. Consequently, the attractive forces between the gas molecules at normal pressure are unimportant. They increase appreciably when the gas is at high pressure as the molecules are then on the average very much closer.

In the bulk of the gas, the resultant force of attraction between a particular molecule and those all round it is zero when averaged over a period. Molecules which strike the wall of the containing vessel, however, are retarded by an unbalanced force due to molecules behind them. The observed pressure p of a gas is thus *less* than the pressure in the ideal case, when the attractive forces due to molecules is zero.

Van der Waals derived an expression for this pressure 'defect'. He considered that it was proportional to the product of the number of molecules per second striking unit area of the wall and the number per unit volume behind them, since this is a measure of the force of attraction. For a given volume of gas, both these numbers are proportional to the *density* of the gas. Consequently the pressure defect, p_1 say, is proportional to $\rho \times \rho$ or ρ^2 . For a fixed mass of gas, $\rho \propto 1/V$, where V is the volume. Thus $p_1 = a/V^2$, where a is a constant for the particular gas. Taking into account the attractive forces between the molecules, it follows that, if p is the observed pressure, the gas pressure in the bulk of the gas $= p + a/V^2$.

The attraction of the walls on the molecules arriving there is to increase their velocity from v say to $v + \Delta v$. Immediately after rebounding from the walls, however, the force of attraction decreases the velocity to v again. Thus the attraction of the walls has no net effect on the momentum change due to collision. Likewise, the increase in momentum of the walls due to their attraction by the molecules arriving is lost after the molecules rebound.

The effect of the volume actually occupied by all the molecules is represented by a constant b , so that the volume of the space in which they move is not V but $(V - b)$. The magnitude of b is not the actual volume of the molecules, as if they were swept into one corner of the space, since they are in constant motion. b has been estimated to be about four times the actual volume.

Surface Tension.

We now consider in detail a phenomenon of a liquid surface called *surface tension*. As we shall soon show, surface tension is due to intermolecular attraction.

It is a well-known fact that some insects, for example a water-carrier, are able to walk across a water surface; that a drop of water may remain suspended for some time from a tap before falling, as if the water particles were held together in a bag; that mercury gathers into small droplets when spilt; and that a dry steel needle may be made, with

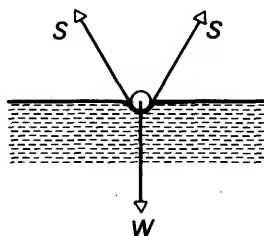


FIG. 5.2 Needle floating on water

care, to float on water, Fig. 5.2. These observations suggest that *the surface of a liquid acts like an elastic skin covering the liquid or is in a state of tension*. Thus forces S in the liquid support the weight W of the needle, as shown in Fig. 5.2.

Energy of Liquid Surface. Molecular theory

The fact that a liquid surface is in a state of tension can be explained by the intermolecular forces discussed on p. 125. In the bulk of the liquid, which begins only a few molecular diameters downwards from the surface, a particular molecule such as A is surrounded by an equal number of molecules on all sides. This can be seen by drawing a sphere round A. Fig. 5.3. The average distance apart of the molecules is such that the attractive forces balance the repulsive forces (p. 145). Thus the average intermolecular force between A and the surrounding molecules is zero. Fig. 5.3.

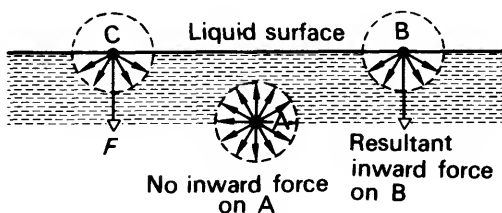


FIG. 5.3 Molecular forces in liquid

Consider now a molecule such as C or B in the surface of the liquid. There are very few molecules on the vapour side above C or B compared with the liquid below, as shown by drawing a sphere round C or B. Thus if C is displaced very slightly upward, a resultant attractive force F on C, due to the large number of molecules below C, now has to be overcome. It follows that if all the molecules in the surface were removed to infinity, a definite amount of work is needed. *Consequently molecules in the surface have potential energy.* A molecule in the bulk of the liquid forms bonds with more neighbours than one in the surface. Thus bonds must be broken, i.e. work must be done, to bring a molecule into the surface. Molecules in the surface of the liquid hence have more potential energy than those in the bulk.

Surface area. Shape of drop

The potential energy of any system in stable equilibrium is a minimum. Thus under surface tension forces, the area of a liquid surface will have the least number of molecules in it, that is, the surface area of a given volume of liquid is a minimum. Mathematically, it can be shown that the shape of a given volume of liquid with a minimum surface area is a *sphere*.

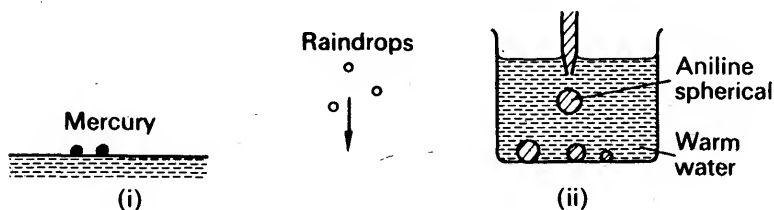


FIG. 5.4 Liquid drops

This is why raindrops, and small droplets of mercury, are approximately spherical in shape. Fig. 5.4 (i). To eliminate completely the effect of gravitational forces, Plateau placed a drop of oil in a mixture of alcohol and water of the same density. In this case the weight of the drop is counterbalanced by the upthrust of the surrounding liquid. He then observed that the drop was a perfect sphere. Plateau's 'spherule' experiment can be carried out by warming water in a beaker and then carefully introducing aniline with the aid of a pipette. Fig. 5.4 (ii). At room temperature the density of aniline is slightly greater than water. At a higher temperature the densities of the two liquids are roughly the same and the aniline is then seen to form *spheres*, which rise and fall in the liquid.

A soap bubble is spherical because its weight is extremely small and the liquid shape is then mainly due to surface tension forces. Although the density of mercury is high, small drops of mercury are spherical. The ratio of surface area ($4\pi r^2$) to weight (or volume, $4\pi r^3/3$) of a sphere is proportional to the ratio r^2/r^3 , or to $1/r$. Thus the smaller

the radius, the greater is the influence of surface tension forces compared to the weight. Large mercury drops, however, are flattened on top. This time the effect of gravity is relatively greater. The shape of the drop conforms to the principle that the sum of the gravitational potential energy and the surface energy must be a minimum, and so the centre of gravity moves down as much as possible.

Lead shot is manufactured by spraying lead from the top of a tall tower. As they fall, the small drops form spheres under the action of surface tension forces.

Surface tension definition. Units, dimensions

Since the surface of a liquid acts like an elastic skin, the surface is in a state of tension. A blown-up football bladder has a surface in a state of tension. This is a very rough analogy because the surface tension of a bladder increases as the surface area increases, whereas the surface tension of a liquid is independent of surface area. Any line in the bladder surface is then acted on by two equal and opposite forces, and if the bladder is cut with a knife the rubber is drawn away from the incision by the two forces present.

R. C. Brown and others have pointed out that molecules in the surface of a liquid have probably a less dense packing than those in the bulk of the liquid, as there are fewer molecules in the surface when its area is a minimum. The average separation between molecules in the surface are then slightly greater than those inside. On average, then, the force between neighbouring molecules in the surface are attractive (see p. 125). This would explain the existence of surface tension.

The *surface tension*, γ , of a liquid, sometimes called the *coefficient of surface tension*, is defined as the *force per unit length acting in the surface at right angles to one side of a line drawn in the surface*. In Fig. 5.5 AB represents a line 1 m long. The unit of γ is *newton metre⁻¹* (N m^{-1}).

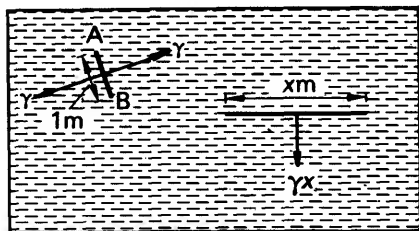


FIG. 5.5 Surface tension

The 'magnitude' of γ depends on the temperature of the liquid and on the medium on the other side of the surface. For water at 20°C in contact with air, $\gamma = 7.26 \times 10^{-2}$ newton metre⁻¹. For mercury at 20°C in contact with air, $\gamma = 46.5 \times 10^{-2}$ N m^{-1} . The surface tension of a water-oil (olive-oil) boundary is 2.06×10^{-2} N m^{-1} , and for a mercury-water boundary it is 42.7×10^{-2} N m^{-1} .

Since surface tension γ is a 'force per unit length', the dimensions of

$$\begin{aligned}\text{surface tension} &= \frac{\text{dimensions of force}}{\text{dimensions of length}} = \frac{MLT^{-2}}{L} \\ &= MT^{-2}.\end{aligned}$$

We shall see later that surface tension can be defined also in terms of surface energy (p. 146).

Some surface tension phenomena

The effect of surface tension forces in a soap film can be demonstrated by placing a thread B carefully on a soap film formed in a metal ring A, Fig. 5.6 (i). The surface tension forces on both sides of the thread counterbalance, as shown in Fig. 5.6 (i). If the film enclosed by the thread is pierced, however, the thread is pulled out into a circle by the surface tension forces F at the junction of the air and soap-film, Fig. 5.6 (ii). Observe that the film has contracted to a minimum area.

Another demonstration of surface tension forces can be made by sprinkling light dust or lycopodium powder over the surface of water contained in a dish. If the middle of the water is touched with the end of a glass rod which had previously been dipped into soap solution, the powder is carried away to the sides by the water. The explanation lies in the fact that the surface tension of water is greater than that of a

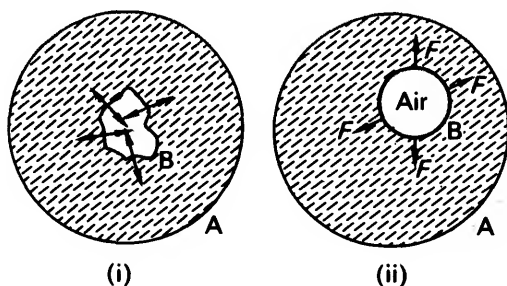


FIG. 5.6 Contraction of surface

soap-film (p. 136). The resultant force at the place where the rod touched the water is hence *away from* the rod, and thus the powder moves away from the centre towards the sides of the vessel.

A toy duck moves by itself across the surface of water when it has a small bag of camphor attached to its base. The camphor lowers the surface tension of the water in contact with it, and the duck is urged across the water by the resultant force on it.

Capillarity

When a capillary tube is immersed in water, and then placed vertically with one end in the liquid, observation shows that the water rises in the tube to a height above the surface. The narrower the tube,

the greater is the height to which the water rises, Fig. 5.7 (i). See also p. 140). This phenomenon is known as *capillarity*, and it occurs when blotting-paper is used to dry ink. The liquid rises up the pores of the paper when it is pressed on the ink.

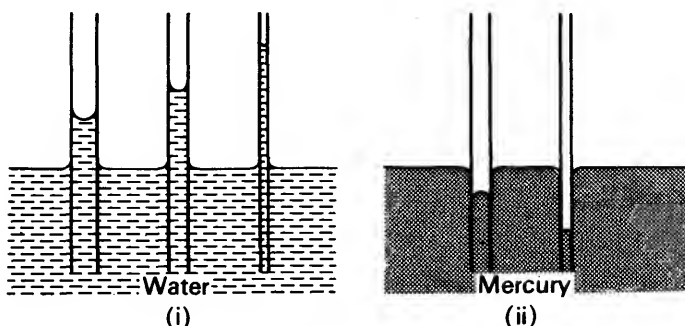


FIG. 5.7 Capillary rise and fall

When a capillary tube is placed inside mercury, however, the liquid is depressed *below* the outside level, Fig. 5.7 (ii). The depression increases as the diameter of the capillary tube decreases. See also p. 141.

Angle of Contact

In the case of water in a glass capillary tube, observation of the meniscus shows that it is hemispherical if the glass is clean, that is, the glass surface is tangential to the meniscus where the water touches it. In other cases where liquids rise in a capillary tube, the tangent BN to the liquid surface where it touches the glass may make an acute angle θ with the glass, Fig. 5.8 (i). The angle θ is known as the *angle of contact* between the liquid and the glass, and is always measured *through the liquid*. The angle of contact between two given surfaces varies largely with their freshness and cleanliness. The angle of contact between water and very clean glass is zero, but when the glass is not clean the angle of contact may be about 8° for example. The angle of contact between alcohol and very clean glass is zero.

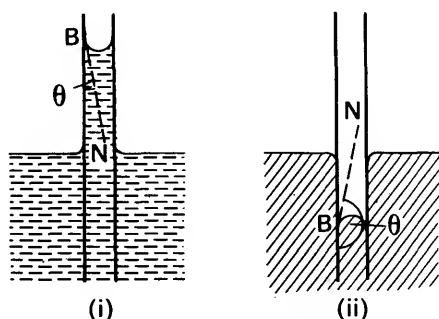


FIG. 5.8 Angle of contact

When a capillary tube is placed inside mercury, observation shows that the surface of the liquid is depressed in the tube and is convex upwards. Fig. 5.8 (ii). The tangent BN to the mercury at the point B where the liquid touches the glass thus makes an obtuse angle, θ , with the glass when measured through the liquid. We shall see later (p. 160) that a liquid will rise in a capillary tube if the angle of contact is acute, and that a liquid will be depressed in the tube if the angle of contact is obtuse. For the same reason, clean water spreads over, or 'wets', a clean glass surface when spilt on it, Fig. 5.9 (i); the angle of contact is zero. On the other hand, mercury gathers itself into small pools or globules when spilt on glass, and does not 'wet' glass, Fig. 5.9 (ii). The angle of contact is obtuse.

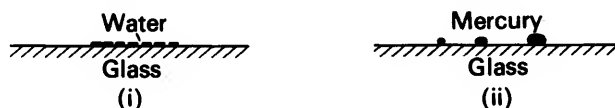


FIG. 5.9 Water and mercury on glass

The difference in behaviour of water and mercury on clean glass can be explained in terms of the attraction between the molecules of these substances. It appears that the force of *cohesion* between two molecules of water is less than the force of *adhesion* between a molecule of water and a molecule of glass; and thus water spreads over glass. On the other hand, the force of cohesion between two molecules of mercury is greater than the force of adhesion between a molecule of mercury and a molecule of glass; and thus mercury gathers in pools when spilt on glass.

Angle of Contact measurement

The angle of contact can be found by means of the method outlined in Fig. 5.10 (i), (ii).

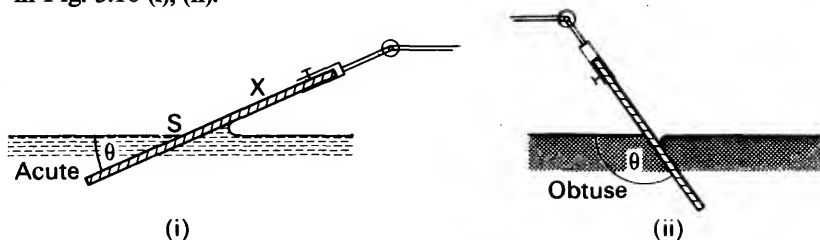


FIG. 5.10 Angle of contact measurement

A plate X of the solid is placed at varying angles to liquid until the surface S appears to be *plane* at X. The angle θ made with the liquid surface is then the angle of contact. For an obtuse angle of contact, a similar method can be adopted. In the case of mercury and glass, a thin plane mirror enables the liquid surface to be seen by reflection. For a freshly-formed mercury drop in contact with a clean glass plate, the angle of contact is 137° .

Measurement of Surface Tension by Capillary Tube Method

Theory. Suppose γ is the magnitude of the surface tension of a liquid such as water, which rises up a clean glass capillary tube and has an angle of contact zero. Fig. 5.11 shows a section of the meniscus M at B, which is a hemisphere. Since the glass AB is a tangent to the liquid, the surface tension forces, which act along the boundary of the liquid with the air, act vertically downwards on the glass. By the law of action and reaction, the glass exerts an equal force in an upward direction on the liquid. Now surface tension, γ , is the force per unit length acting in the surface of the liquid, and the length of liquid in contact with the glass is $2\pi r$, where r is the radius of the capillary tube.

$$\therefore 2\pi r \times \gamma = \text{upward force on liquid} \quad (1)$$

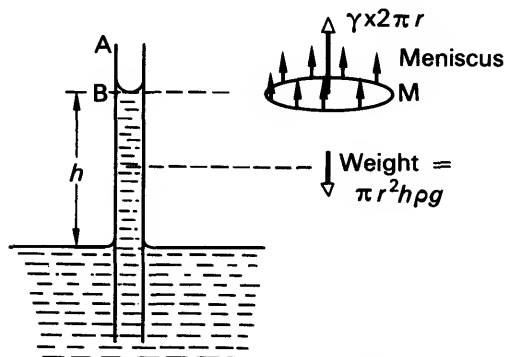


FIG. 5.11 Rise in capillary tube—theory

If γ is in newton metre $^{-1}$ and r is in metres, then the upward force is in newtons.

This force supports the weight of a column of height h above the outside level of liquid. The volume of the liquid = $\pi r^2 h$, and thus the mass, m , of the liquid column = volume \times density = $\pi r^2 h \rho$, where ρ is the density. The weight of the liquid = $mg = \pi r^2 h \rho g$.

If ρ is in kg m^{-3} , r and h in metres, and $g = 9.8 \text{ m s}^{-2}$, then $\pi r^2 h \rho g$ is in newtons.

From (1), it now follows that

$$\begin{aligned} \therefore 2\pi r \gamma &= \pi r^2 h \rho g \\ \therefore \gamma &= \frac{r h \rho g}{2} \end{aligned} \quad (2)$$

If $r = 0.2 \text{ mm} = 0.2 \times 10^{-3} \text{ m}$, $h = 6.6 \text{ cm}$ for water = $6.6 \times 10^{-2} \text{ m}$, and $\rho = 1 \text{ g cm}^{-3} = 1000 \text{ kg m}^{-3}$, then

$$\gamma = \frac{0.2 \times 10^{-3} \times 6.6 \times 10^{-2} \times 1000 \times 9.8}{2} = 6.5 \times 10^{-2} \text{ N m}^{-1}$$

In deriving this formula for γ it should be noted that we have (i) assumed the glass to be a tangent to the liquid surface meeting it, (ii) neglected the weight of the small amount of liquid above the bottom of the meniscus at B, Fig. 5.11.

Experiment. In the experiment, the capillary tube C is supported in a beaker Y, and a pin P, bent at right angles at two places, is attached to C by a rubber band. P is adjusted until its point just touches the horizontal level of the liquid in the beaker. A travelling microscope is now focussed on to the meniscus M in C, and then it is focussed on to the point of P, the beaker being removed for this observation. In this way the height h of M above the level in the beaker is determined. The

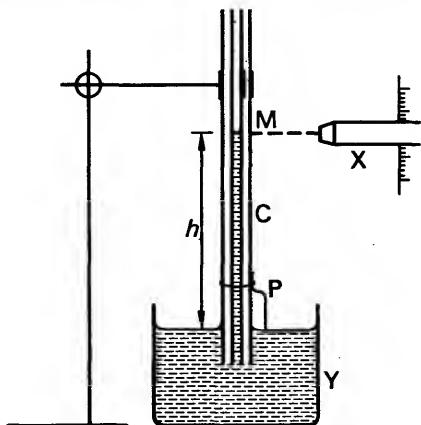


FIG. 5.12 Surface tension by capillary rise

radius of the capillary at M can be found by cutting the tube at this place and measuring the diameter by the travelling microscope; or by measuring the length, l , and mass, m , of a mercury thread drawn into the tube, and calculating the radius, r , from the relation $r = \sqrt{m/\pi l \rho}$, where ρ is the density of mercury. The surface tension γ is then calculated from the formula $\gamma = rh\rho g/2$. Its magnitude for water at 15°C is 7.33×10^{-2} newton metre $^{-1}$.

Measurement of Surface Tension by Microscope Slide

Besides the capillary tube method, the surface tension of water can be measured by weighing a microscope slide in air, and then lowering it until it just meets the surface of water, Fig. 5.13. The surface tension force acts vertically downward round the boundary of the slide, and pulls the slide down. If a and b are the length and thickness of the slide,

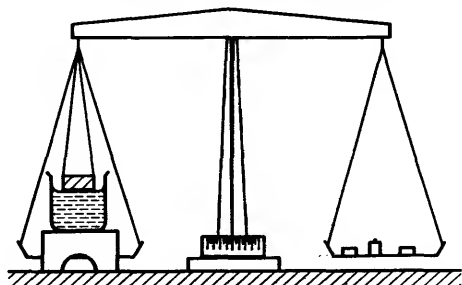


FIG. 5.13 Surface tension by microscope slide

then, since γ is the force per unit length in the liquid surface and $(2a+2b)$ is the length of the boundary of the slide, the downward force $= \gamma(2a+2b)$. If the mass required to counterbalance the force is m , then

$$\gamma(2a+2b) = mg,$$

$$\therefore \gamma = \frac{mg}{2a+2b}.$$

If $m = 0.88$ gramme, $a = 6.0$ cm, $b = 0.2$ cm, then:

$$\gamma = \frac{0.88 \times 10^{-3} \text{ (kg)} \times 9.8 \text{ (m s}^{-2}\text{)}}{2 \times (6 + 0.2) \times 10^{-2} \text{ (m)}} = 7.0 \times 10^{-2} \text{ N m}^{-1}.$$

Surface Tension of a Soap Solution

The surface tension of a soap solution can be found by a similar method. A soap-film is formed in a three-sided metal frame ABCD, and the apparent weight is found, Fig. 5.14. When the film is broken by piercing it, the decrease in the apparent weight, mg , is equal to the

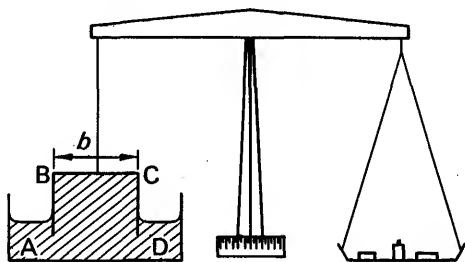


FIG. 5.14 Surface tension of soap film

surface tension force acting downwards when the film existed. This is equal to $2\gamma b$, where $b = BC$, since the film has *two* sides.

$$\therefore 2\gamma b = mg,$$

$$\therefore \gamma = \frac{mg}{2b}.$$

It will be noted that the surface tension forces on the sides AB, CD of the frame act horizontally, and their resultant is zero.

A soap film can be supported in a vertical rectangular frame but a film of water can not. This is due to the fact that the soap drains downward in a vertical film, so that the top of the film has a lower concentration of soap than the bottom. The surface tension at the top is thus *greater* than at the bottom (soap diminishes the surface tension of pure water). The upward pull on the film by the top bar is hence greater than the downward pull on the film by the lower bar. The net upward pull supports the weight of the film. In the case of pure water, however, the surface tension would be the same at the top and bottom, and hence there is no net force in this case to support a water film in a rectangular frame.

Pressure Difference in a Bubble or Curved Liquid Surface

As we shall see presently, the magnitude of the curvature of a liquid, or of a bubble formed in a liquid, is related to the surface tension of the liquid.

Consider a bubble formed inside a liquid, Fig. 5.15. If we consider the equilibrium of *one half*, B, of the bubble, we can see that the surface tension force on B plus the force on B due to the external pressure p_1 = the force on B due to the internal pressure p_2 inside the bubble. The force on B due to the pressure p_1 is given by $\pi r^2 \times p_1$, since πr^2 is the area of the circular face of B and pressure is 'force per unit area'; the force on B due to the pressure p_2

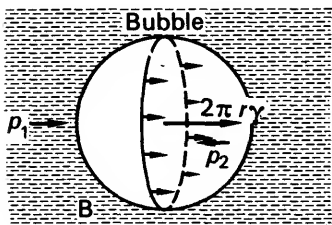


FIG. 5.15 Excess pressure in bubble

is given similarly by $\pi r^2 \times p_2$. The surface tension force acts round the *circumference* of the bubble, which has a length $2\pi r$; thus the force is $2\pi r\gamma$. It follows that

$$2\pi r\gamma + \pi r^2 p_1 = \pi r^2 p_2.$$

Simplifying,

$$\therefore 2\gamma = r(p_2 - p_1),$$

or

$$p_2 - p_1 = \frac{2\gamma}{r}.$$

Now $(p_2 - p_1)$ is the excess pressure, p , in the bubble over the outside pressure.

$$\therefore \text{excess pressure, } p, = \frac{2\gamma}{r} \quad \dots \quad (1)$$

Although we considered a bubble, the same formula for the excess pressure holds for any curved liquid surface or meniscus, where r is its radius of curvature and γ is its surface tension, provided the angle of contact is zero. If the angle of contact is θ , the formula is modified by replacing γ by $\gamma \cos \theta$. Thus, in general,

$$\text{excess pressure, } p, = \frac{2\gamma \cos \theta}{r} \quad \dots \quad (2)$$

Excess Pressure in Soap Bubble

A soap bubble has two liquid surfaces in contact with air, one inside the bubble and the other outside the bubble. The force on one half, B, of the bubble due to surface tension forces is thus $\gamma \times 2\pi r \times 2$, i.e., $\gamma \times 4\pi r$, Fig. 5.16. For the equilibrium of B, it follows that

$$4\pi r\gamma + \pi r^2 p_1 = \pi r^2 p_2,$$

where p_2, p_1 are the pressures inside and outside the bubble respectively. Simplifying,

$$\therefore p_2 - p_1 = \frac{4\gamma}{r},$$

$$\therefore \text{excess pressure } p = \frac{4\gamma}{r} \quad \dots \quad (3)$$

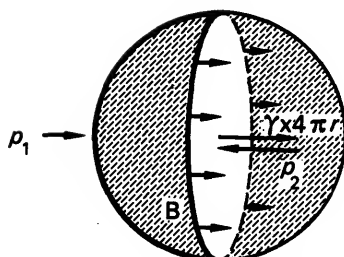


FIG. 5.16 Bubble in water

This result for excess pressure should be compared with the result obtained for a bubble formed inside a liquid, equation (1).

If γ for a soap solution is $25 \times 10^{-3} \text{ N m}^{-1}$, the excess pressure inside a bubble of radius 0.5 cm or $0.5 \times 10^{-2} \text{ m}$ is hence given by:

$$p = \frac{4 \times 25 \times 10^{-3}}{0.5 \times 10^{-2}} = 20 \text{ N m}^{-2}.$$

Two soap-bubbles of unequal size can be blown on the ends of a tube, communication between them being prevented by a closed tap in the middle. If the tap is opened, the *smaller* bubble is observed to collapse gradually and the size of the larger bubble increases. This can be explained from our formula $p = 4\gamma/r$, which shows that the pressure of air inside the smaller bubble is greater than that inside the larger bubble. Consequently air flows from the smaller to the larger bubble when communication is made between the bubbles, and the smaller bubble thus gradually collapses.

Since the excess pressure in a bubble is inversely-proportional to the radius, the pressure needed to form a very small bubble is high. This explains why one needs to blow hard to start a balloon growing. Once the balloon has grown, less air pressure is needed to make it expand more.

Surface Tension of Soap-Bubble

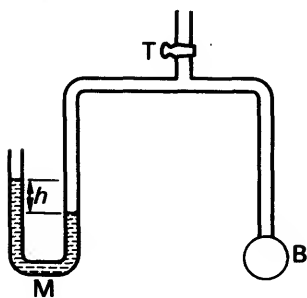


FIG. 5.17

Surface tension of soap-bubble

The surface tension of a soap solution can be measured by blowing a small soap-bubble at the end B of a tube connected to a manometer M, Fig. 5.17. The tap T is then closed, the diameter d of the bubble is measured by a travelling microscope, and the difference in levels h of the liquid in the manometer is observed with the same instrument. The excess pressure, p , in the bubble $= h\rho g$, where ρ is the density of the liquid in M.

$$\therefore h\rho g = \frac{4\gamma}{r} = \frac{4\gamma}{d/2}.$$

$$\therefore \gamma = \frac{h\rho g d}{8}.$$

Rise or Fall of Liquids in Capillary Tubes

From our knowledge of the angle of contact and the excess pressure on one side of a curved liquid surface, we can deduce that some liquids will rise in a capillary tube, whereas others will be depressed.

Suppose the tube A is placed in water, for example, Fig. 5.18 (i). At first the liquid surface becomes concave upwards in the tube, because the angle of contact with the glass is zero. Consequently the pressure on the air side, X, of the curved surface is greater than the pressure on the liquid side Y by $2\gamma/r$, where γ is the surface tension and r is the radius of curvature of the tube. But the pressure at X is atmospheric, H . Hence the pressure at Y must be less than atmospheric by $2\gamma/r$. Fig. 5.18 (i) is therefore impossible because it shows the pressure at Y equal to the atmospheric pressure. Thus, as shown in Fig. 5.18 (ii), the liquid ascends the tube to a height h such that the pressure at N is less than at M by $2\gamma/r$, Fig. 5.18 (ii). A similar argument shows that a liquid rises in a capillary tube when the angle of contact is acute.

The angle of contact between mercury and glass is obtuse (p. 133). Thus when a capillary tube is placed in mercury the liquid first curves downwards. The pressure inside the liquid just below the curved surface is now greater than the pressure on the other side, which is atmospheric, and the mercury therefore moves down the tube until the excess pressure $= 2\gamma \cos \theta/r$, with the usual notation. A liquid thus falls in a capillary tube if the angle of contact is obtuse.

Capillary Rise and Fall by Pressure Method

We shall now calculate the capillary rise of water by the excess pressure formula $p = 2\gamma/r$, or $p = 2\gamma \cos \theta/r$.

In the case of a capillary tube dipping into water, the angle of contact is practically zero, Fig. 5.19 (i). Thus if p_2 is the pressure of the atmosphere, and p_1 is the pressure in the liquid, we have

$$p_2 - p_1 = \frac{2\gamma}{r}.$$

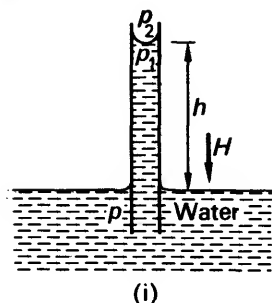


FIG. 5.19
Excess
pressure
application

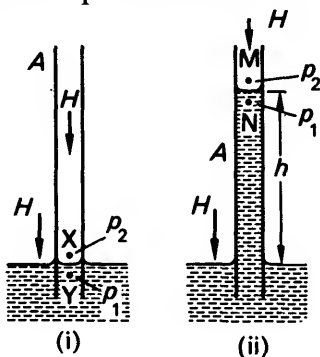
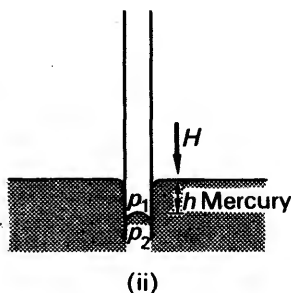


FIG. 5.18
Capillary rise by excess pressure



Now if H is the atmospheric pressure, h is the height of the liquid in the tube and ρ its density,

$$p_2 = H \text{ and } p_1 = H - h\rho g,$$

$$\therefore H - (H - h\rho g) = \frac{2\gamma}{r},$$

$$\therefore h\rho g = \frac{2\gamma}{r},$$

$$\therefore h = \frac{2\gamma}{r\rho g} \quad \text{. (i)}$$

The formula shows that h increases as r decreases, i.e., the narrower the tube, the greater is the height to which the water rises (see Fig. 5.7 (i), p. 132).

If the height l of the tube above the water is *less* than the calculated value of h in the above formula, the water surface at the top of the tube now meets it at an *acute angle of contact* θ . The radius of the meniscus is therefore $r/\cos \theta$, and $l\rho g = 2\gamma/(r/\cos \theta)$, or

$$l = \frac{2\gamma \cos \theta}{r\rho g} \quad \text{. (ii)}$$

Dividing (ii) by (i), it follows that

$$\cos \theta = \frac{l}{h}.$$

Thus suppose water rises to a height of 10 cm in a capillary tube when it is placed in a beaker of water. If the tube is pushed down until the top is only 5 cm above the outside water surface, then $\cos \theta = \frac{5}{10} = 0.5$. Thus $\theta = 60^\circ$. The meniscus now makes an angle of contact of 60° with the glass. As the tube is pushed down further, the angle of contact increases beyond 60° . When the top of the tube is level with the water in the beaker, the meniscus in the tube becomes plane. (See Example 2, p. 141.)

With Mercury in Glass

Suppose that the depression of the mercury inside a tube of radius r is h , Fig. 5.19 (ii). The pressure p_2 below the curved surface of the mercury is then greater than the (atmospheric) pressure p_1 outside the curved surface; and, from our general result,

$$p_2 - p_1 = \frac{2\gamma \cos \theta}{r},$$

where θ is the supplement of the obtuse angle of contact of mercury with glass, that is, θ is an acute angle and its cosine is positive. But $p_1 = H$ and $p_2 = H + h\rho g$, where H is the atmospheric pressure.

$$\therefore (H + h\rho g) - H = \frac{2\gamma \cos \theta}{r}.$$

$$\therefore h\rho g = \frac{2\gamma \cos \theta}{r}$$

$$\therefore h = \frac{2\gamma \cos \theta}{r\rho g} \quad (1)$$

The height of depression, h , thus increases as the radius r of the tube decreases. See Fig. 5.7 (ii), p. 132.

EXAMPLES

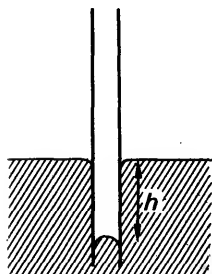


FIG. 5.20 Example

1. Define surface tension of a liquid and describe a method of finding this quantity for alcohol.

It water rises in a capillary tube 5.8 cm above the free surface of the outer liquid, what will happen to the mercury level in the same tube when it is placed in a dish of mercury? Illustrate this by the aid of a diagram. Calculate the difference in level between the mercury surfaces inside the tube and outside. (S.T. of water = $75 \times 10^{-3} \text{ N m}^{-1}$. S.T. of mercury = $547 \times 10^{-3} \text{ N m}^{-1}$. Angle of contact of mercury with clean glass = 130° . Density of mercury = 13600 kg m^{-3} .) (L.)

Second part. The mercury is depressed a distance h below the outside level, and is convex upward, Fig. 5.20. Suppose r is the capillary tube radius.

For water, $h = 5.8 \text{ cm} = 5.8 \times 10^{-2} \text{ m}$, $\gamma = 75 \times 10^{-3} \text{ newton m}^{-1}$, $\rho = 1000 \text{ kg m}^{-3}$, $g = 9.8 \text{ m s}^{-2}$.

From $\gamma = rh\rho g/2$,

$$\therefore 75 \times 10^{-3} = r \times 5.8 \times 10^{-2} \times 1000 \times 9.8/2 \quad (r \text{ in metre}).$$

For mercury, $\rho = 13.6 \times 10^3 \text{ kg m}^{-3}$, $\gamma = 547 \times 10^{-3} \text{ newton m}^{-1}$.

$$\begin{aligned} \therefore h &= \frac{2\gamma \cos 50^\circ}{r\rho g} \\ &= \frac{2 \times 547 \times 10^{-3} \cos 50^\circ \times 5.8 \times 10^{-2} \times 1000 \times 9.8}{13.6 \times 10^3 \times 9.8 \times 75 \times 10^{-3} \times 2} \\ &= 0.02 \text{ m} = 2 \text{ cm}. \end{aligned}$$

2. On what grounds would you anticipate some connection between the surface tension of a liquid and its latent heat of vaporization?

A vertical capillary tube 10 cm long tapers uniformly from an internal diameter of 1 mm at the lower end to 0.5 mm at the upper end. The lower end is just touching the surface of a pool of liquid of surface tension $6 \times 10^{-2} \text{ N m}^{-1}$, density 1200 kg m^{-3} and zero angle of contact with the tube. Calculate the capillary rise, justifying your method. Explain what will happen to the meniscus if the tube is slowly lowered vertically until the upper end is level with the surface of the pool. (O. & C.)

Suppose S is the meniscus at a height h cm above the liquid surface. The tube tapers uniformly and the change in radius for a height of 10 cm is $(0.05 - 0.025)$ or 0.025 cm , so that the change in radius per cm height is 0.0025 cm . Thus at a height h cm, radius of meniscus S is given by

$$r = (0.05 - 0.0025 h) \times 10^{-2} \text{ m}$$

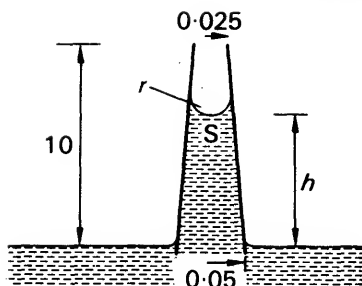


FIG. 5.21 Example

The pressure above S is atmospheric, A . The pressure below S is $(A - h\rho g)$.

$$\therefore \text{pressure difference} = (h \times 10^{-2})\rho g = \frac{2\gamma}{r} = \frac{200\gamma}{0.05 - 0.0025h}$$

$$\therefore 0.05h - 0.0025h^2 = \frac{200\gamma}{\rho g} = \frac{200 \times 6 \times 10^{-2}}{10^{-2} \times 1200 \times 9.8} = 0.102$$

$$\therefore h^2 - 20h = -40 \text{ (approx.)}$$

$$\therefore (h - 10)^2 = 100 - 40 = 60$$

$$\therefore h = 10 - \sqrt{60} = 2.2 \text{ cm.}$$

If the tube is slowly lowered the meniscus reaches the top at some stage. On further lowering the tube the angle of contact changes from zero to an acute angle. When the upper end is level with liquid surface the meniscus becomes plane.

3. 'The surface tension of water is 7.5×10^{-2} newton m^{-1} and the angle of contact of water with glass is zero.' Explain what these statements mean. Describe an experiment to determine *either* (a) the surface tension of water, or (b) the angle of contact between paraffin wax and water.

A glass U-tube is inverted with the open ends of the straight limbs, of diameters respectively 0.500 mm and 1.00 mm, below the surface of water in a beaker. The air pressure in the upper part is increased until the meniscus in one limb is level with the water outside. Find the height of water in the other limb. (The density of water may be taken as 1000 kg m^{-3}). (L.)

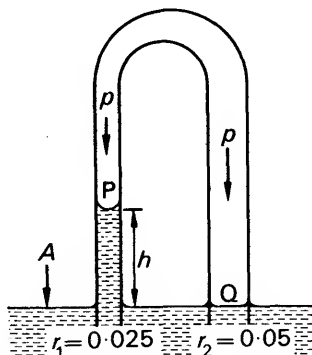


FIG. 5.22 Example (radii in cm)

Suppose p is the air pressure inside the U-tube when the meniscus Q is level with the water outside and P is the other meniscus at a height h . Let A be the atmospheric pressure. Then, if r_1 is the radius at P,

$$p - (A - h\rho g) = \frac{2\gamma}{r_1} \quad (i)$$

since the pressure in the liquid below P is $(A - h\rho g)$.

The pressure in the liquid below Q = A . Hence, for Q,

$$p - A = \frac{2\gamma}{r_2} \quad (ii)$$

where r_2 is the radius.

From (i) and (ii), it follows that

$$\begin{aligned} h\rho g &= \frac{2\gamma}{r_1} - \frac{2\gamma}{r_2} \\ \therefore h &= \frac{1}{\rho g} \left[\frac{2\gamma}{r_1} - \frac{2\gamma}{r_2} \right] \\ &= \frac{1}{9800} \left[\frac{2 \times 0.075}{0.25 \times 10^{-3}} - \frac{2 \times 0.075}{0.5 \times 10^{-3}} \right] \\ &= 3.1 \times 10^{-2} \text{ m (approx).} \end{aligned}$$

Effects of surface tension in measurements

When a hydrometer is used to measure relative density or density, the surface tension produces a downward force F on the hydrometer. If r is the radius of the stem and the angle of contact is zero, then $F = 2\pi r\gamma$. For a narrow stem, the error produced in reading the relative density from the graduations is small.

Another case of an undesirable surface tension effect occurs in measurements of the height of liquid columns in glass tubes. The height of mercury in a barometer, for example, is depressed by surface tension (p. 114). If the tubes are wide, surface tension forces can be neglected. If they are narrow, the forces must be taken into account. As an illustration, consider an inverted U-tube dipping into two liquids B and C. Fig. 5.23. These can be drawn up into the tubes to heights

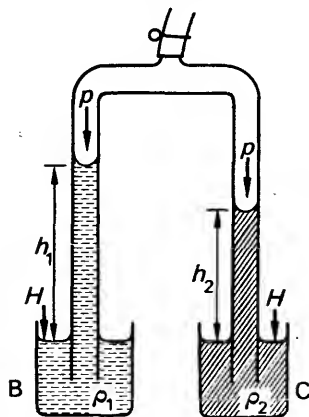


FIG. 5.23 Comparison of densities

h_1 , h_2 respectively above the outside level. In the absence of surface tension forces, $p + h_1\rho_1g =$ atmospheric pressure H , where p is the air pressure at the top of the tubes $= p + h_2\rho_2g$. Thus $h_1\rho_1 = h_2\rho_2$, or $h_1/h_2 = \rho_2/\rho_1$. Thus the liquid densities may be compared from the ratio of the heights of the liquid columns.

To take account of surface tension, we proceed as follows. Using the notation on p. 137.

$$p - p_1 = \frac{2\gamma_1}{r_1},$$

where p is the air pressure at the top of the tubes, p_1 is the pressure in the liquid near the meniscus of the tube in B, γ_1 is the surface tension of the liquid, and r_1 is the radius. But, from hydrostatics, $p_1 = H - h_1\rho_1g$.

$$\begin{aligned}\therefore p - (H - h_1\rho_1g) &= \frac{2\gamma_1}{r_1} \\ \therefore H - p &= h_1\rho_1g - \frac{2\gamma_1}{r_1} \quad \quad \quad (i)\end{aligned}$$

If γ_2 is the surface tension of the liquid in C, and r_2 is the radius of the tube in the liquid, then, by similar reasoning,

$$H - p = h_2\rho_2g - \frac{2\gamma_2}{r_2} \quad \quad \quad (ii)$$

From (i) and (ii),

$$\therefore h_2\rho_2g - \frac{2\gamma_2}{r_2} = h_1\rho_1g - \frac{2\gamma_1}{r_1}.$$

Re-arranging,

$$\therefore h_2 = \frac{\rho_1}{\rho_2}h_1 - \frac{2}{\rho_2g}\left(\frac{\gamma_1}{r_1} - \frac{\gamma_2}{r_2}\right),$$

which is an equation of the form $y = mx + c$, where c is a constant, $h_2 = y$, $h_1 = x$, and $\rho_1/\rho_2 = m$. Thus by taking different values of h_2 and h_1 , and plotting h_2 against h_1 , a straight-line graph is obtained whose slope is equal to ρ_1/ρ_2 , the ratio of the densities. In this way the effect of the surface tension can be eliminated.

Variation of Surface Tension with Temperature. Jaeger's Method

By forming a bubble inside a liquid, and measuring the excess pressure, JAEGER was able to determine the variation of the surface tension of a liquid with temperature. One form of the apparatus is shown in Fig. 5.24 (i). A capillary or drawn-out tubing A is connected to a vessel W containing a funnel C, so that air is driven slowly through A when water enters W through C, so that air is driven slowly through A when water enters W through C. The capillary A is placed inside a beaker containing the liquid L, and a bubble forms slowly at the end of A when air is passed through it at a slow rate.

Fig. 5.24. (ii) shows the bubble at three possible stages of growth. The radius grows from that at a to a hemispherical shape at b . Here its pressure is larger since the radius is smaller. If we consider the bubble growing to c , the radius of c would be greater than that of b and hence it cannot contain the increasing pressure. The downward force on the bubble due to the pressure, in fact, would be greater than

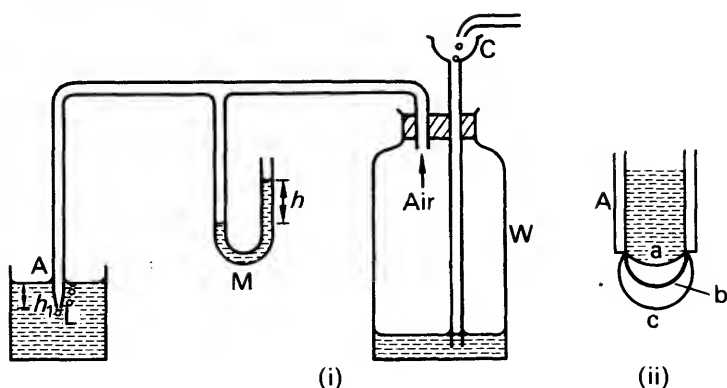


FIG. 5.24 Jaeger's method

the upward force due to surface tension. Hence the bubble becomes unstable and breaks away from A when its radius is the same as that of A. Thus as the bubble grows the pressure in it increases to a maximum, and then decreases as the bubble breaks away. The maximum pressure is observed from a manometer M containing a light oil of density ρ , and a series of observations are taken as several bubbles grow.

The maximum pressure inside the bubble $= H + h\rho g$ where h is the maximum difference in levels in the manometer M, and H is the atmospheric pressure. The pressure outside the bubble $= H + h_1\rho_1 g$, where h_1 is the depth of the orifice of A below the level of the liquid L, and ρ_1 is the latter's density.

$$\therefore \text{excess pressure} = (H + h\rho g) - (H + h_1\rho_1 g) = h\rho g - h_1\rho_1 g.$$

$$\text{But} \quad \text{excess pressure} = \frac{2\gamma}{r},$$

where r is the radius of the orifice of A (p. 158).

$$\therefore \frac{2\gamma}{r} = h\rho g - h_1\rho_1 g,$$

$$\therefore \gamma = \frac{rg}{2}(h\rho - h_1\rho_1).$$

By adding warm liquid to the vessel containing L, the variation of the surface tension with temperature can be determined. Experiment shows that the surface tension of liquids, and water in particular, decreases with increasing temperature along a fairly smooth curve. Various formulae relating the surface tension to temperature have been proposed, but none has been found to be completely satisfactory. The decrease of surface tension with temperature may be attributed to the greater average separation of the molecules at higher temperature. The force of attraction between molecules is then reduced, and hence the surface energy is reduced, as can be seen from the potential energy curve on p. 126.

Surface Tension and Surface Energy

We now consider the surface energy of a liquid and its relation to its surface tension γ . Consider a film of liquid stretched across a horizontal frame ABCD, Fig. 5.25. Since γ is the force per unit length, the force on the rod BC of length $l = \gamma \times 2l$, because there are two surfaces to the film.

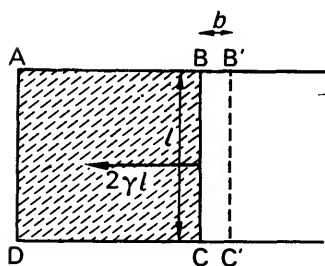


FIG. 5.25
Surface energy and work

Suppose the rod is now moved a distance b from BC to B'C' against the surface tension forces, so that the surface area of the film increases. The temperature of the film then usually decreases, in which case the surface tension alters (p. 145). If the surface area increases under *isothermal* (constant temperature) conditions, however, the surface tension is constant; and we can then say that, if γ is the surface tension at that temperature,

$$\begin{aligned}\text{work done in enlarging surface area} &= \text{force} \times \text{distance}, \\ &= 2\gamma l \times b = \gamma \times 2lb.\end{aligned}$$

But $2lb$ is the total increase in surface area of the film.

$$\therefore \text{work done per unit area in enlarging area} = \gamma.$$

Thus the surface tension, γ , can be defined as *the work done per unit area in increasing the surface area of a liquid under isothermal conditions*. This is also called the *free surface energy*.

Surface energy and Latent heat

Inside a liquid molecules move about in all directions, continually breaking and reforming bonds with neighbours. If a molecule in the surface passes into the vapour outside, a definite amount of energy is needed to permanently break the bonds with molecules in the liquid. This amount of energy is the work done in overcoming the inward force on a molecule in the surface, discussed on p. 129. Thus the energy needed to evaporate a liquid is related to its surface energy or surface tension. The latent heat of vaporisation, which is the energy needed to change liquid to vapour at the boiling point, is therefore related to surface energy.

Surface energy

As we have seen, when the surface area of a liquid is increased, the surface energy is increased. The molecules which then reach the surface are slowed up by the inward force, so the average translational kinetic energy of all the liquid molecules is reduced. On this account the liquid cools while the surface is increased, and heat flows in from the surroundings to restore the temperature.

The increase in the *total surface energy per unit area* E is thus given by

$$E = \gamma + H \quad (1)$$

where H is the heat per unit area from the surroundings. Advanced theory shows that $H = -\theta \left(\frac{d\gamma}{d\theta} \right)$, where θ is the absolute temperature and $d\gamma/d\theta$ is the corresponding gradient of the γ v. θ graph, the variation of surface tension with temperature. Thus

$$E = \gamma - \theta \frac{d\gamma}{d\theta} \quad (2)$$

In practice, since γ decreases with rising temperature, $d\gamma/d\theta$ is negative, and E is thus greater than γ . At 15°C , for example, $\gamma = 74 \times 10^{-3} \text{ N m}^{-1}$, $d\gamma/d\theta = -0.15 \times 10^{-3} \text{ N m}^{-1} \text{ K}^{-1}$, $\theta = 288 \text{ K}$. Thus, from (2),

$$E = (74 + 288 \times 0.15) \times 10^{-3} = 0.117 \text{ N m}^{-1} = 0.117 \text{ J m}^{-2}.$$

The variation of E with temperature is shown in Fig. 5.26, together with the similar variation of L , the latent heat of vaporisation (see p. 146). Both vanish at the critical temperature, since no liquid exists above the critical temperature whatever the pressure.

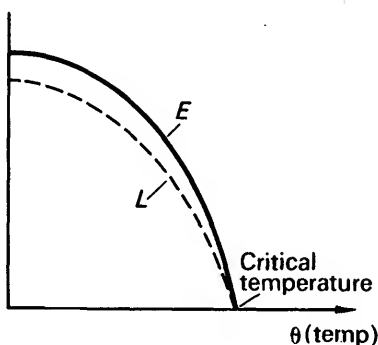


FIG. 5.26 Variation of E and L with temperature

EXAMPLES

1. A soap bubble in a vacuum has a radius of 3 cm and another soap bubble in the vacuum has a radius of 6 cm. If the two bubbles coalesce under isothermal conditions, calculate the radius of the bubble formed.

Since the bubbles coalesce under isothermal conditions, the surface tension γ is constant. Suppose R is the radius in cm, $R \times 10^{-2} \text{ m}$, of the bubble formed.

Then $\text{work done} = \gamma \times \text{surface area} = \gamma \times 8\pi R^2 \times 10^{-4}$

But $\text{original work done} = (\gamma \times 8\pi \cdot 3^2 + \gamma \times 8\pi \cdot 6^2) \times 10^{-4}$

$$\therefore \gamma \times 8\pi R^2 = \gamma \times 8\pi \cdot 3^2 + \gamma \cdot 8\pi \cdot 6^2.$$

$$\therefore R^2 = 3^2 + 6^2.$$

$$\therefore R = \sqrt{3^2 + 6^2} = 6.7 \text{ cm}.$$

2. (i) Calculate the work done against surface tension forces in blowing a soap bubble of 1 cm diameter if the surface tension of soap solution is $2.5 \times 10^{-2} \text{ N m}^{-1}$. (ii) Find the work required to break up a drop of water of radius 0.5 cm into drops of water each of radii 1 mm. (Surface tension of water = $7 \times 10^{-2} \text{ N m}^{-1}$.)

(i) The original surface area of the bubble is zero, and the final surface area = $2 \times 4\pi r^2$ (two surfaces of bubble) = $(2 \times 4\pi \times 0.5^2) \times 10^{-4} = 2\pi \times 10^{-4} \text{ m}^2$.

\therefore work done = $\gamma \times$ increase in surface area.

$$= 2.5 \times 10^{-2} \times 2\pi \times 10^{-4} = 1.57 \times 10^{-5} \text{ J.}$$

(ii) Since volume of a drop = $\frac{4}{3}\pi r^3$,

$$\text{number of drops formed} = \frac{\frac{4}{3}\pi \times 0.5^3}{\frac{4}{3}\pi \times 0.1^3} = 125.$$

\therefore final total surface area of drops

$$= 125 \times 4\pi r^2 = 125 \times 4\pi \times 0.1^2 \times 10^{-4},$$

$$= 5\pi \times 10^{-4} \text{ m}^2.$$

But original surface area of drop = $4\pi \times 0.5^2 \times 10^{-4} = \pi \times 10^{-4} \text{ m}^2$.

\therefore work done = $\gamma \times$ change in surface area,

$$= 7 \times 10^{-2} \times (5\pi - \pi) \times 10^{-4} = 8.8 \times 10^{-5} \text{ J.}$$

EXERCISES 5

What are the missing words in the statements 1–8?

1. The units of surface tension are ...
2. The dimensions of surface tension are ...
3. Small drops of mercury are spherical because the surface area is a ...
4. The excess pressure in a soap-bubble is given by ...
5. The excess pressure at the meniscus of water in a capillary tube is ...
6. A liquid will not 'wet' the surface of a solid if the angle of contact is ...
7. Surface tension may be defined as the 'force ...'
8. Surface tension may also be defined as the '... per unit area'.

Which of the following answers, A, B, C, D or E, do you consider is the correct one in the statements 9–12?

9. A molecule of a liquid which reaches the surface from the interior gains energy because *A* it reaches the surface with higher speed than when inside the liquid, *B* it overcomes a force of repulsion on molecules at the surface, *C* it overcomes a force of attraction on molecules at the surface, *D* its temperature increases, *E* the gravitational potential energy due to the earth is then higher.

10. If a section of a soap bubble through its centre is considered, the force on one half due to surface tension is *A* $2\pi r\gamma$, *B* $4\pi r\gamma$, *C* $\pi r^2\gamma$, *D* $2\gamma/r$, *E* $2\pi r^2\gamma$.

11. If water has a surface tension of $7 \times 10^{-2} \text{ N m}^{-1}$ and an angle of contact with water of zero, it rises in a capillary of diameter 0.5 mm to a height of *A* 70 cm, *B* 7.0 cm, *C* 6.2 cm, *D* 5.7 cm, *E* 0.5 cm.

12. In an experiment to measure the surface tension of a liquid by rise in a capillary tube which tapers, the necessary radius r would be best obtained *A* by cutting the tube at the position of the meniscus and measuring the diameter here directly, *B* by drawing up a thread of mercury of length l and using 'mass =

$\pi^2 l \rho'$, C by measuring the diameter of the lower end of the tube with a travelling microscope, D by measuring the upper end of the tube with a travelling microscope, E by finding the average of the two measurements in C and D .

13. Define *surface tension*. A rectangular plate of dimensions 6 cm by 4 cm and thickness 2 mm is placed with its largest face flat on the surface of water. Calculate the force due to surface tension on the plate. What is the downward force due to surface tension if the plate is placed vertical and its longest side just touches the water? (Surface tension of water = $7.0 \times 10^{-2} \text{ N m}^{-1}$.)

14. What are the *dimensions* of surface tension? A capillary tube of 0.4 mm diameter is placed vertically inside (i) water of surface tension $6.5 \times 10^{-2} \text{ N m}^{-1}$ and zero angle of contact, (ii) a liquid of density 800 kg m^{-3} , surface tension $5.0 \times 10^{-2} \text{ N m}^{-1}$ and angle of contact 30° . Calculate the height to which the liquid rises in the capillary in each case.

15. Define the *angle of contact*. What do you know about the angle of contact of a liquid which (i) wets glass, (ii) does not wet glass?

A capillary tube is immersed in water of surface tension $7.0 \times 10^{-2} \text{ N m}^{-1}$ and rises 6.2 cm. By what depth will mercury be depressed if the same capillary is immersed in it? (Surface tension of mercury = 0.54 N m^{-1} ; angle of contact between mercury and glass = 140° ; density of mercury = 13600 kg m^{-3} .)

16. (i) A soap-bubble has a diameter of 4 mm. Calculate the pressure inside it if the atmospheric pressure is 10^5 N m^{-2} . (Surface tension of soap solution = $2.8 \times 10^{-2} \text{ N m}^{-1}$.) (ii) Calculate the radius of a bubble formed in water if the pressure outside it is $1.000 \times 10^5 \text{ N m}^{-2}$ and the pressure inside it is $1.001 \times 10^5 \text{ N m}^{-2}$. (Surface tension of water = $7.0 \times 10^{-2} \text{ N m}^{-1}$.)

17. Define *surface tension* of a liquid. State the units in which it is usually expressed and give its dimensions in mass, length, and time.

Derive an expression for the difference between the pressure inside and outside a spherical soap bubble. Describe a method of determining surface tension, based on the difference of pressure on the two sides of a curved liquid surface or film. (L .)

18. Explain briefly (a) the approximately spherical shape of a rain drop, (b) the movement of tiny particles of camphor on water, (c) the possibility of floating a needle on water, (d) why a column of water will remain in an open vertical capillary tube after the lower end has been dipped in water and withdrawn. (N .)

19. Define the terms surface tension, angle of contact. Describe a method for measuring the surface tension of a liquid which wets glass. List the principal sources of error and state what steps you would take to minimize them.

A glass tube whose inside diameter is 1 mm is dipped vertically into a vessel containing mercury with its lower end 1 cm below the surface. To what height will the mercury rise in the tube if the air pressure inside it is $3 \times 10^3 \text{ N m}^{-2}$ below atmospheric pressure? Describe the effect of allowing the pressure in the tube to increase gradually to atmospheric pressure. (Surface tension of mercury = 0.5 N m^{-1} , angle of contact with glass = 180° , density of mercury = 13600 kg m^{-3} , $g = 9.81 \text{ m s}^{-2}$.) ($O. \& C.$)

20. Explain how to measure the surface tension of a soap film.

The diameters of the arms of a U-tube are respectively 1 cm and 1 mm. A liquid of surface tension $7.0 \times 10^{-2} \text{ N m}^{-1}$ is poured into the tube which is placed vertically. Find the difference in levels in the two arms. The density may be taken as 1000 kg m^{-3} and the contact angle zero. (L .)

21. Explain what is meant by surface tension, and show how its existence is accounted for by molecular theory.

Find an expression for the excess pressure inside a soap-bubble of radius R and surface tension T . Hence find the work done by the pressure in increasing the radius of the bubble from a to b . Find also the increase in surface area of the bubble, and in the light of this discuss the significance of your result. (C.)

22. A clean glass capillary tube, of internal diameter 0.04 cm, is held vertically with its lower end below the surface of clean water in a beaker, and with 10 cm of the tube above the surface. To what height will the water rise in the tube? What will happen if the tube is now depressed until only 5 cm of its length is above the surface? The surface tension of water is $7.2 \times 10^{-2} \text{ N m}^{-1}$.

Describe, and give the theory of some method, other than that of the rise in a capillary tube, of measuring surface tension. (O. & C.)

23. Explain (a) in terms of molecular forces why the water is drawn up above the horizontal liquid level round a steel needle which is held vertically and partly immersed in water, (b) why, in certain circumstances, a steel needle will rest on a water surface. In each case show the relevant forces on a diagram. (N.)

24. The force between two molecules may be regarded as an attractive force which increases as their separation decreases and a repulsive force which is only important at small separations and which there varies very rapidly. Draw sketch graphs (a) for force-separation, (b) for potential-energy separation. On each graph mark the equilibrium distance and on (b) indicate the energy which would be needed to separate two molecules initially at the equilibrium distance.

With the help of your graphs discuss briefly the resulting motion if the molecules are displaced from the equilibrium position. (N.)

25. Explain briefly the meaning of *surface tension* and *angle of contact*.

Account for the following: (a) A small needle may be placed on the surface of water in a beaker so that it 'floats', and (b) if a small quantity of detergent is added to the water the needle sinks.

A solid glass cylinder of length l , radius r and density σ is suspended with its axis vertical from one arm of a balance so that it is partly immersed in a liquid of density ρ . The surface tension of the liquid is γ and its angle of contact with the glass is α . If W_1 is the weight required to achieve a balance when the cylinder is in air and W_2 is the weight required to balance the cylinder when it is partly immersed with a length h ($< l$) below the free surface of the liquid, derive an expression for the value of $W_1 - W_2$. If this method were used to measure the surface tension of a liquid, why would the result probably be less accurate than that obtained from a similar experiment using a thin glass plate? (O. & C.)

26. Explain in terms of molecular forces why some liquids spread over a solid surface whilst others do not.

A glass capillary tube of uniform bore of diameter 0.050 cm is held vertically with its lower end in water. Calculate the capillary rise. Describe and explain what happens if the tube is lowered so that 4.0 cm protrudes above the water surface. Assume that the surface tension of water is $7.0 \times 10^{-2} \text{ N m}^{-1}$. (N.)

27. Define *surface tension*. Describe how the surface tension of water at room temperature may be determined by using a capillary tube. Derive the formula used to calculate the result.

A hydrometer has a cylindrical glass stem of diameter 0.50 cm. It floats in water of density 1000 kg m^{-3} and surface tension $7.2 \times 10^{-2} \text{ N m}^{-1}$. A drop of

liquid detergent added to the water reduces the surface tension to $5.0 \times 10^{-2} \text{ N m}^{-1}$. What will be the change in length of the exposed portion of the glass stem? Assume that the relevant angle of contact is always zero. (N.)

28. The lower end of a vertical clean glass capillary tube is just immersed in water. Why does water rise up the tube?

A vertical capillary tube of internal radius $r \text{ m}$ has its lower end dipping in water of surface tension $T \text{ newton m}^{-1}$. Assuming the angle of contact between water and glass to be zero, obtain from first principles an expression for the pressure excess which must be applied to the upper end of the tube in order just to keep the water levels inside and outside the tube the same.

A capillary of internal diameter 0.7 mm is set upright in a beaker of water with one end below the surface; air is forced slowly through the tube from the upper end, which is also connected to a U-tube manometer containing a liquid of density 800 kg m^{-3} . The difference in levels on the manometer is found to build up to 9.1 cm , drop to 4.0 cm , build up to 9.1 cm again, and so on. Estimate (a) the depth of the open end of the capillary below the free surface of the water in the beaker, (b) the surface tension of water. [State clearly any assumptions you have made in arriving at these estimates.] (O.)

29. It is sometimes stated that, in virtue of its surface tension, the surface of a liquid behaves as if it were a stretched rubber membrane. To what extent do you think this analogy is justified?

Explain why the pressure inside a spherical soap bubble is greater than that outside. How would you investigate experimentally the relation between the excess pressure and the radius of the bubble? Show on a sketch graph the form of the variation you would expect to obtain.

If olive oil is sprayed on to the surface of a beaker of hot water, it remains as separated droplets on the water surface; as the water cools, the oil forms a continuous thin film on the surface. Suggest a reason for this phenomenon. (C.)

30. Describe the capillary tube method of measuring the surface tension of a liquid.

An inverted U-tube (Hare's apparatus) for measuring the specific gravity of a liquid was constructed of glass tubing of internal diameter about 2 mm . The following observations of the heights of balanced columns of water and another liquid were obtained:

Height of water (cm)	2.8	4.2	5.4	6.9	8.5	9.8	11.6
Height of liquid (cm)	2.0	3.8	5.3	7.0	9.1	10.7	13.0

Plot the above results, explain why the graph does not pass through the origin, and deduce from the graph an accurate value for the specific gravity of the liquid. (N.)

31. How does simple molecular theory account for surface tension? Illustrate your account by explaining the rise of water up a glass capillary.

A light wire frame in the form of a square of side 5 cm hangs vertically in water with one side in the water-surface. What additional force is necessary to pull the frame clear of the water? Explain why, if the experiment is performed with soap-solution, as the force is increased a vertical film is formed, whereas with pure water no such effect occurs. (Surface tension of water is $7.4 \times 10^{-2} \text{ N m}^{-1}$.) (O. & C.)

32. Define *surface tension* and state the effect on the surface tension of water of raising its temperature.

Describe an experiment to measure the surface tension of water over the range of temperatures from 20°C to 70°C . Why is the usual capillary rise method unsuitable for this purpose?

Two unequal soap bubbles are formed one on each end of a tube closed in the middle by a tap. State and explain what happens when the tap is opened to put the two bubbles into connection. Give a diagram showing the bubbles when equilibrium has been reached. (L.)

chapter six

Elasticity

Elasticity

A bridge, when used by traffic during the day, is subjected to loads of varying magnitude. Before a steel bridge is erected, therefore, samples of the steel are sent to a research laboratory, where they undergo tests to find out whether the steel can withstand the loads to which it is likely to be subjected.

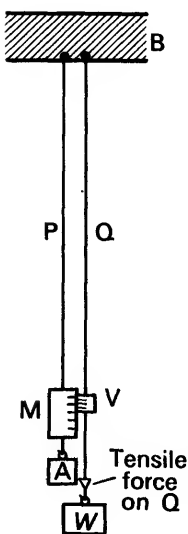


FIG. 6.1
Tensile force

Fig. 6.1 illustrates a simple laboratory method of discovering useful information about the property of steel we are discussing. Two long thin steel wires, P, Q, are suspended beside each other from a rigid support B, such as a girder at the top of the ceiling. The wire P is kept taut by a weight A attached to its end and carries a scale M graduated in centimetres. The wire Q carries a vernier scale V which is alongside the scale M.

When a load W such as 1 kgf is attached to the end of Q, the wire increases in length by an amount which can be read from the change in the reading on the vernier V. If the load is taken off and the reading on V returns to its original value, the wire is said to be **elastic** for loads from zero to 1 kgf, a term adopted by analogy with an elastic thread. When the load W is increased to 2 kgf the extension (increase in length) is obtained from V again; and if the reading on V returns to origin

value when the load is removed the wire is said to be elastic at least for loads from zero to 2 kgf.

The extension of a thin wire such as Q for increasing loads may be found by experiments to be as follows:

W (kgf)	0	1	2	3	4	5	6	7	8
Extension (mm.)	0	0.14	0.28	0.42	0.56	0.70	0.85	1.01	1.19

Elastic Limit

When the extension, e , is plotted against the load, W , a graph is obtained which is a *straight line* OA, followed by a curve ABY rising slowly

at first and then very sharply, Fig. 6.2. Up to about 5 kgf, then, the results in the table show that the extension increased by 0.14 mm for every kgf which is added to the wire. Further, the wire returned to its

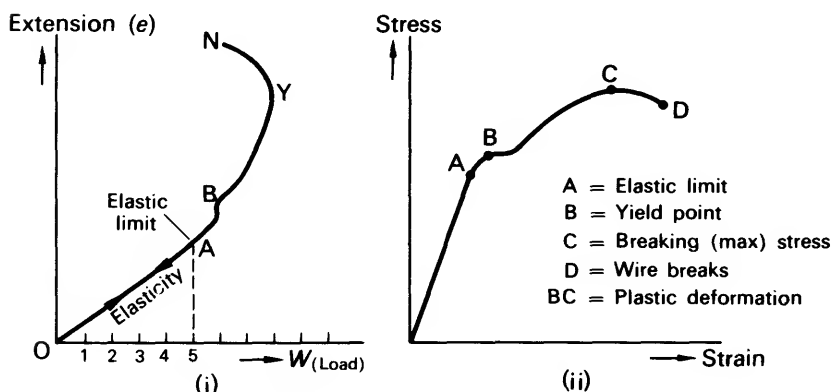


FIG. 6.2 Extension v. Load

original length when the load was removed. For loads greater than about 5 kgf, however, the extension increases relatively more and more, and the wire now no longer returns to its original length when it is unloaded. The wire is thus permanently strained, and A corresponds to its *elastic limit*.

Hooke's Law

From the straight line graph OA, we deduce that *the extension is proportional to the load or tension in the wire when the elastic limit is not exceeded*. This is known as *Hooke's law*, after ROBERT HOOKE, founder of the Royal Society, who discovered the relation in 1676. The law shows that when a molecule of a solid is displaced farther from its mean position, the restoring force is proportional to its displacement (see p. 126). One may therefore conclude that the molecules of a solid are undergoing simple harmonic motion (p. 44).

The measurements also show that it would be dangerous to load the wire with weights greater than 5 kilogrammes, the elastic limit, because the wire then suffers a permanent strain. Similar experiments in the research laboratory enable scientists to find the maximum load which a steel bridge, for example, should carry for safety. Rubber samples are also subjected to similar experiments, to find the maximum safe tension in rubber belts used in machinery.

Yield Point. Ductile and Brittle Substances. Breaking Stress

Careful experiments show that, for mild steel and iron for example, the molecules of the wire begin to 'slide' across each other soon after the load exceeds the elastic limit, that is, the material becomes *plastic*. This is indicated by the slight 'kink' at B beyond A in Fig. 6.2 (i), and it is called the *yield point* of the wire. The change from an elastic to

a plastic stage is shown by a sudden increase in the extension, and as the load is increased further the extension increases rapidly along the curve YN and the wire then snaps. The *breaking stress* of the wire is the corresponding force per unit area of cross-section of the wire. Substances such as those just described, which elongate considerably and undergo plastic deformation until they break, are known as *ductile* substances. Lead, copper and wrought iron are ductile. Other substances, however, break just after the elastic limit is reached; they are known as *brittle* substances. Glass and high carbon steels are brittle.

Brass, bronze, and many alloys appear to have no yield point. These materials increase in length beyond the elastic limit as the load is increased without the appearance of a plastic stage.

The strength and ductility of a metal, its ability to flow, are dependent on defects in the metal crystal lattice. Such defects may consist of a missing atom at a site or a *dislocation* at a plane of atoms. Plastic deformation is the result of the 'slip' of atomic planes. The latter is due to the movement of dislocations, which spreads across the crystal.

Tensile Stress and Tensile Strain. Young's Modulus

We have now to consider the technical terms used in the subject of elasticity of wires. When a force or tension F is applied to the end of a wire of cross-sectional area A , Fig. 6.3,

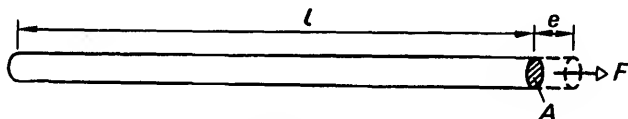


FIG. 6.3 Tensile stress and tensile strain

$$\text{the tensile stress} = \text{force per unit area} = \frac{F}{A} \quad (1)$$

If the extension of the wire is e , and its original length is l ,

$$\text{the tensile strain} = \text{extension per unit length} = \frac{e}{l} \quad (2)$$

Suppose 2 kg is attached to the end of a wire of length 2 metres of diameter 0.64 mm, and the extension is 0.60 mm. Then

$$F = 2 \text{ kgf} = 2 \times 9.8 \text{ N}, \quad A = \pi \times 0.032^2 \text{ cm}^2 = \pi \times 0.032^2 \times 10^{-4} \text{ m}^2.$$

$$\therefore \text{tensile stress} = \frac{2 \times 9.8}{\pi \times 0.032^2 \times 10^{-4}} \text{ N m}^{-2},$$

$$\text{and} \quad \text{tensile strain} = \frac{0.6 \times 10^{-3} \text{ metre}}{2 \text{ metre}} = 0.3 \times 10^{-3}.$$

It will be noted that 'stress' has units such as 'newton m^{-2} '; 'strain' has no units because it is the ratio of two lengths.

A *modulus of elasticity* of the wire, called **Young's modulus** (E), is defined as the ratio

$$E = \frac{\text{tensile stress}}{\text{tensile strain}} \quad (3)$$

Thus

$$E = \frac{F/A}{e/l}$$

Using the above figures,

$$\begin{aligned} E &= \frac{2 \times 9.8 / (\pi \times 0.032^2 \times 10^{-4})}{0.3 \times 10^{-3}}, \\ &= \frac{2 \times 9.8}{\pi \times 0.032^2 \times 10^{-4} \times 0.3 \times 10^{-3}}, \\ &= 2.0 \times 10^{11} \text{ N m}^{-2}. \end{aligned}$$

It should be noted that Young's modulus, E , is calculated from the ratio stress:strain only when the wire is under 'elastic' conditions, that is, the load does not then exceed the elastic limit (p.154). Fig. 6.2 (ii) shows the general stress-strain diagram for a ductile material.

Dimensions of Young's Modulus

As stated before, the 'strain' of a wire has no dimensions of mass, length, or time, since, by definition, it is the ratio of two lengths. Now

$$\begin{aligned} \text{dimensions of stress} &= \frac{\text{dimensions of force}}{\text{dimensions of area}} \\ &= \frac{MLT^{-2}}{L^2} \\ &= ML^{-1}T^{-2}. \end{aligned}$$

\therefore dimensions of Young's modulus, E ,

$$\begin{aligned} &= \frac{\text{dimensions of stress}}{\text{dimensions of strain}} \\ &= ML^{-1}T^{-2}. \end{aligned}$$

Determination of Young's Modulus

The magnitude of Young's modulus for a material in the form of a wire can be found with the apparatus illustrated in Fig. 6.1, p.153, to which the reader should now refer. The following practical points should be specially noted:

(1) The wire is made *thin* so that a moderate load of several kilograms produces a large tensile stress. The wire is also made *long* so that a measurable extension is produced.

(2) The use of two wires, P, Q, of the same material and length, eliminates the correction for (i) the yielding of the support when loads are added to Q, (ii) changes of temperature.

(3) Both wires should be free of kinks, otherwise the increase in length cannot be accurately measured. The wires are straightened by attaching weights to their ends, as shown in Fig. 6.1.

(4) A vernier scale is necessary to measure the extension of the wire since this is always small. The 'original length' of the wire is measured from the top B to the vernier V by a ruler, since an error of 1 millimetre is negligible compared with an original length of several metres. For very accurate work, the extension can be measured by using a spirit level between the two wires, and adjusting a vernier screw to restore the spirit level to its original reading after a load is added.

(5) The diameter of the wire must be found by a micrometer screw gauge at several places, and the average value then calculated. The area of cross-section, $A = \pi r^2$, where r is the radius.

(6) The readings on the vernier are also taken when the load is gradually removed in steps of 1 kilogramme; they should be very nearly the same as the readings on the vernier when the weights were added, showing that the elastic limit was not exceeded. Suppose the reading on V for loads, W , of 1 to 6 kilogramme are a, b, c, d, e, f , as follows:

W (kgf)	1	2	3	4	5	6
Reading on V	a	b	c	d	e	f

The average extension for 3 kilogramme is found by taking the average of $(d - a)$, $(e - b)$, and $(f - c)$. Young's modulus can then be calculated from the relation stress/strain, where the stress $= 3 \times 9.8/\pi r^2$, and the strain $=$ average extension/original length of wire (p. 155).

Magnitude of Young's Modulus

Mild steel (0.2% carbon) has a Young's modulus value of about 2.0×10^{11} newton m^{-2} , copper has a value about 1.2×10^{11} newton m^{-2} ; and brass a value about 1.0×10^{11} newton m^{-2} .

The breaking stress (tenacity) of cast-iron metal is about 1.5×10^8 newton m^{-2} ; the breaking stress of mild steel metal is about 4.5×10^8 newton m^{-2} .

At Royal Ordnance and other Ministry of Supply factories, tensile testing is carried out by placing a sample of the material in a machine known as an *extensometer*, which applies stresses of increasing value along the length of the sample and automatically measures the slight increase in length. When the elastic limit is reached, the pointer on the dial of the machine flickers, and soon after the yield point is reached the sample becomes thin at some point and then breaks. A graph showing the load v. extension is recorded automatically by a moving pen while the sample is undergoing test.

EXAMPLE

Find the maximum load in kgf which may be placed on a steel wire of diameter 0.10 cm if the permitted strain must not exceed $\frac{1}{1000}$ and Young's modulus for steel is 2.0×10^{11} N m^{-2} .

We have $\frac{\text{max. stress}}{\text{max. strain}} = 2 \times 10^{11}$.

$$\therefore \text{max. stress} = \frac{1}{1000} \times 2 \times 10^{11} = 2 \times 10^8 \text{ N m}^{-2}.$$

This is the amount of energy stored in the wire. The formula $\frac{1}{2}Fe$ gives the energy in joule when F is in newton and e is in metre.

Further, since $F = EAe/l$,

$$\text{energy} = \frac{1}{2}EA\frac{e^2}{l}.$$

As an illustration, suppose $E = 2.0 \times 10^{11}$ newton m^{-2} , $A = 3 \times 10^{-2} \text{ cm}^2 = 3 \times 10^{-6} \text{ m}^2$, $e = 1 \text{ mm} = 1 \times 10^{-3} \text{ m}$, $l = 400 \text{ cm} = 4 \text{ m}$. Then

$$\begin{aligned}\text{energy stored} &= \frac{1}{2}EA\frac{e^2}{l} = \frac{1}{2} \times \frac{2 \times 10^{11} \times 3 \times 10^{-6} \times (1 \times 10^{-3})^2}{4} \text{ joule,} \\ &= 0.075 \text{ J.}\end{aligned}$$

The volume of the wire $= Al$. Thus, from (1),

$$\text{energy per unit volume} = \frac{1}{2}\frac{Fe}{Al} = \frac{1}{2}\frac{F}{A} \times \frac{e}{l}.$$

But $F/A = \text{stress}$, $e/l = \text{strain}$,

$$\therefore \text{energy per unit volume} = \frac{1}{2} \text{ stress} \times \text{strain} \quad (2)$$

Graph of F v. e and energy

The energy stored in the wire when it is stretched can also be found from the graph of F v. e . Fig. 6.4. Suppose the wire extension is e_1 when a load F_1 is applied, and the extension increases to e_2 when the load increases to F_2 . If F is the load between F_1 and F_2 at some stage, and Δx is the small extension which then occurs, then

$$\text{energy stored} = \text{work done} = F \cdot \Delta x.$$

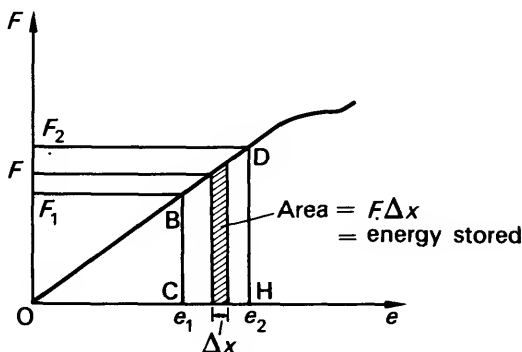


FIG. 6.4 Energy in stretched wire

Now $F \cdot \Delta x$ is represented by the small area between the axis of e and the graph, shown shaded in Fig. 6.4. Thus the total work done between e_1 and e_2 is represented by the area $CBDH$.

If the extension occurs on the straight part of the curve, when Hooke's law is obeyed, then CBDH is a trapezium. The area of a trapezium = half the sum of the parallel sides \times perpendicular distance between them = $\frac{1}{2}(BC + DH) \times CH = \frac{1}{2}(F_1 + F_2)(e_2 - e_1)$.

\therefore energy stored = average force \times increase in length.

If the extension occurs beyond the elastic limit, for example, along the curved part of the graph in Fig. 6.4, the energy expended can be obtained from the area between the curve and the axis of e .

EXAMPLES

1. A 20 kg weight is suspended from a length of copper wire 1 mm in radius. If the wire breaks suddenly, does its temperature increase or decrease? Calculate the change in temperature; Young's modulus for copper = $12 \times 10^{10} \text{ N m}^{-2}$; density of copper = 9000 kg m^{-3} ; specific heat capacity of copper = $0.42 \text{ J g}^{-1} \text{ K}^{-1}$. (C.S.)

When the wire is stretched, it gains potential energy equal to the work done on it. When the wire is suddenly broken, this potential energy is released as the molecules return to their original position. The energy is converted into heat and thus the temperature rises.

$$\begin{aligned} \text{Gain in potential energy of molecules} &= \text{work done in stretching wire} \\ &= \frac{1}{2} \text{ force } (F) \times \text{extension } (e). \end{aligned}$$

With the usual notation, $F = EA \frac{e}{l}$

$$\therefore e = \frac{Fl}{E \cdot A} = \frac{(20 \times 9.8) \times l}{12 \times 10^{10} \times \pi \times (10^{-3})^2} \text{ m} = 5.2 \times 10^{-4} l \text{ m},$$

$$\therefore \text{potential energy gained} = \frac{1}{2} \times 20 \times 9.8 \times 5.2 \times 10^{-4} l = 5.1 \times 10^{-2} l \text{ J}$$

Heat capacity of wire = mass \times specific heat capacity

$$= \pi \times (10^{-3})^2 \times 9000 l \times (0.42 \times 1000) = 11.9 l \text{ J K}^{-1}$$

$$\begin{aligned} \therefore \text{temperature rise} &= \frac{\text{potential energy}}{\text{heat capacity}} = \frac{5.1 \times 10^{-2} l}{11.9 l} \\ &= 4.3 \times 10^{-3} \text{ deg C.} \end{aligned}$$

2. Define *stress* and *strain*. Describe the behaviour of a copper wire when it is subjected to an increasing longitudinal stress. Draw a stress-strain diagram and mark on it the elastic region, yield point and breaking stress.

A wire of length 5 m, of uniform circular cross-section of radius 1 mm is extended by 1.5 mm when subjected to a uniform tension of 100 newton. Calculate from first principles the strain energy per unit volume assuming that deformation obeys Hooke's law.

Show how the stress-strain diagram may be used to calculate the work done in producing a given strain, when the material is stretched beyond the Hooke's law region. (O. & C.)

$$\text{Strain energy} = \frac{1}{2} \text{ tension} \times \text{extension}$$

$$\text{Tension} = 100 \text{ newton. Extension} = 1.5 \times 10^{-3} \text{ m.}$$

$$\therefore \text{energy} = \frac{1}{2} \times 100 \times 1.5 \times 10^{-3} = 0.075 \text{ J.}$$

$$\text{Volume of wire} = \text{length} \times \text{area} = 5 \times \pi \times 1 \times 10^{-6} \text{ m}^3.$$

$$\therefore \text{energy per unit volume} = \frac{0.075}{5 \times \pi \times 1 \times 10^{-6}} = 4.7 \times 10^3 \text{ J m}^{-3} \text{ (approx.)}$$

Bulk Modulus

When a gas or a liquid is subjected to an increased pressure the substance contracts. A change in bulk thus occurs, and the *bulk strain* is defined by:

$$\text{strain} = \frac{\text{change in volume}}{\text{original volume}}$$

The *bulk stress* on the substance is the increased force per unit area, by definition, and the bulk modulus, K , is given by:

$$\begin{aligned} K &= \frac{\text{bulk stress}}{\text{bulk strain}} \\ &= \frac{\text{increase in force per unit area}}{\text{change in volume/original volume}} \end{aligned}$$

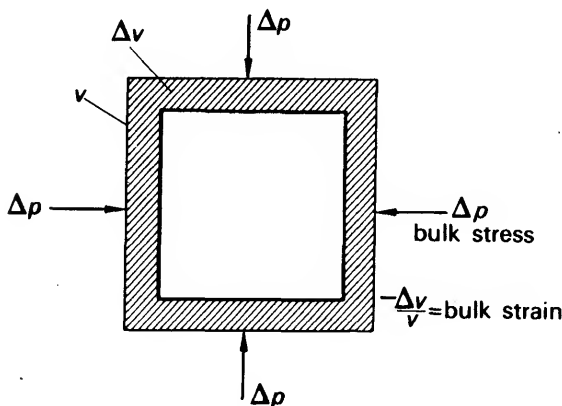


FIG. 6.5 Bulk stress and bulk strain

If the original volume of the substance is v , the change in volume may be denoted by $-\Delta v$ when the pressure increases by a small amount Δp ; the minus indicates that the volume decreases. Thus (Fig. 6.5)

$$K = -\frac{\Delta p}{\Delta v/v}$$

When δp and δv become very small, then, in the limit,

$$K = -v \frac{dp}{dv} \quad (1)$$

The bulk modulus of water is about $2 \times 10^9 \text{ N m}^{-2}$ for pressures in the range 1–25 atmospheres; the bulk modulus of mercury is about

$27 \times 10^9 \text{ N m}^{-2}$. The bulk modulus of gases depends on the pressure, as now explained. Generally, since the volume change is relatively large, the bulk modulus of a gas is low compared with that of a liquid.

Bulk Modulus of a Gas

If the pressure, p , and volume, v , of a gas change under conditions such that

$$pv = \text{constant},$$

which is Boyle's law, the changes are said to be *isothermal* ones. In this case, by differentiating the product pv with respect to v , we have

$$p + v \frac{dp}{dv} = 0.$$

$$\therefore p = -v \frac{dp}{dv}.$$

But the bulk modulus, K , of the gas is equal to $-v \frac{dp}{dv}$ by definition (see p. 161).

$$\therefore K = p \quad \dots \dots \dots (2)$$

Thus the *isothermal bulk modulus is equal to the pressure*.

When the pressure, p , and volume, v , of a gas change under conditions such that

$$pv^\gamma = \text{constant},$$

where $\gamma = c_p/c_v$ = the ratio of the specific heat capacities of the gas, the changes are said to be *adiabatic* ones. This equation is the one obeyed by local values of pressure and volume in air when a sound wave travels through it. Differentiating both sides with respect to v ,

$$\therefore p \times \gamma v^{\gamma-1} + v^\gamma \frac{dp}{dv} = 0,$$

$$\therefore \gamma p = -v \frac{dp}{dv},$$

$$\therefore \text{adiabatic bulk modulus} = \gamma p \quad \dots \dots \dots (3)$$

For air at normal pressure, $K = 10^5 \text{ newton m}^{-2}$ isothermally and $1.4 \times 10^5 \text{ newton m}^{-2}$ adiabatically. The values of K are of the order 10^5 times smaller than liquids as gases are much more compressible.

Velocity of Sound

The velocity of sound waves through any material depends on (i) its density ρ , (ii) its modulus of elasticity, E . Thus if V is the velocity, we may say that

$$V = kE^x \rho^y \quad \dots \dots \dots (i),$$

where k is a constant and x, y are indices we can find by the theory of dimensions (p. 34).

The units of velocity, V , are LT^{-1} ; the units of density ρ are ML^{-3} ; and the units of modulus of elasticity, E , are $\text{ML}^{-1}\text{T}^{-2}$ (see p. 156). Equating the dimensions on both sides of (i),

$$\therefore \text{LT}^{-1} = (\text{ML}^{-1}\text{T}^{-2})^x \times (\text{ML}^{-3})^y.$$

Equating the indices of M, L, T on both sides, we have

$$\begin{aligned}0 &= x + y, \\1 &= -x - 3y, \\-1 &= -2x.\end{aligned}$$

Solving, we find $x = \frac{1}{2}$, $y = -\frac{1}{2}$. Thus $V = kE^{\frac{1}{2}}\rho^{-\frac{1}{2}}$. A rigid investigation shows $k = 1$, and thus

$$V = E^{\frac{1}{2}}\rho^{-\frac{1}{2}} = \sqrt{\frac{E}{\rho}}.$$

In the case of a solid, E is Young's modulus. In the case of air and other gases, and of liquids, E is replaced by the bulk modulus K . Laplace showed that the adiabatic bulk modulus must be used in the case of a gas, and since this is γp , the velocity of sound in a gas is given by the expression

$$V = \sqrt{\frac{\gamma p}{\rho}}.$$

Modulus of Rigidity

So far we have considered the strain in one direction, or tensile strain, to which Young's modulus is applicable and the strain in bulk or volume, to which the bulk modulus is applicable.

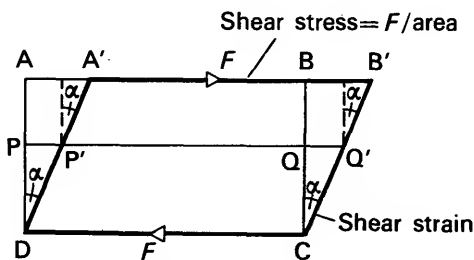


FIG. 6.6 Shear stress and shear strain

Consider a block of material ABCD, such as pitch or plastic for convenience. Fig. 6.6. Suppose the lower plane CD is fixed, and a stress parallel to CD is applied by a force F to the upper side AB. The block then changes its shape and takes up a position A'B'CD. It can now be seen that planes in the material parallel to DC are displaced relative to each other. The plane AB, for example, which was originally directly opposite the plane PQ, is displaced to A'B' and PQ is displaced to P'Q'. The *angular displacement* α is defined as the *shear strain*. α is the angular displacement between any two planes, for example, between CD and P'Q'.

No volume change occurs in Fig. 6.6. Further, since the force along CD is F in magnitude, it forms a *couple* with the force F applied to the upper side AB. The *shear stress* is defined as the 'shear force per unit area' on the face AB (or CD), as in Young's modulus or the bulk

modulus. Unlike the case for these moduli, however, the shear stress has a turning or 'displacement' effect owing to the couple present. The solid does not collapse because in a strained equilibrium position such as A'B'CD in Fig. 6.6, the external couple acting on the solid due to the forces F is balanced by an opposing couple due to stresses inside the material.

If the elastic limit is not exceeded when a shear stress is applied, that is, the solid recovers its original shape when the stress is removed, the *modulus of rigidity*, G , is defined by:

$$G = \frac{\text{shear stress (force per unit area)}}{\text{shear strain (angular displacement, } \alpha)}$$

Shear strain has no units; shear stress has units of newton m^{-2} . The modulus of rigidity of copper is $4.8 \times 10^{10} \text{ N m}^{-2}$; for phosphor-bronze it is $4.4 \times 10^{10} \text{ N m}^{-2}$, and for quartz fibre it is $3.0 \times 10^{10} \text{ N m}^{-2}$.

If a spiral spring is stretched, all parts of the spiral become twisted. The applied force has thus developed a 'torsional' or shear strain. The extension of the spring hence depends on its modulus of rigidity, in addition to its dimensions.

Torsion wire

In sensitive current-measuring instruments, a very weak control is needed for the rotation of the instrument coil. This may be provided by using a long elastic or *torsion wire* of phosphor bronze in place of a spring. The coil is suspended from the lower end of the wire and when it rotates through an angle θ , the wire sets up a weak opposing couple equal to $c\theta$, where c is the elastic constant of the wire. Quartz fibres are very fine but comparatively strong, and have elastic properties. They are also used for sensitive control (see p. 61).

The magnitude of c , the elastic constant, can be derived as follows. Consider a wire of radius a , length l , modulus of rigidity G , fixed at the upper end and twisted by a couple of moment C at the other end. If we take a section of the cylindrical wire between radii r and $r + \delta r$, then a 'slice' of the material ODBX has been sheared through an angle α to a position ODB₁X where X is the centre of the lower end of the wire. Fig. 6.7. From the definition of modulus of rigidity, $G = \text{torsional stress} \div \text{torsional strain} = F/A \div \alpha$, where F is the tangential force applied over an area A .

Now $A = \text{area of circular annulus at lower end} = 2\pi r \cdot \delta r$.

$$\therefore F = G\alpha A = G \cdot 2\pi r \cdot \delta r \cdot \alpha$$

From Fig. 6.7, it follows that $BB_1 = l\alpha$, and $BB_1 = r\theta$.

$$\therefore l\alpha = r\theta, \text{ or } \alpha = r\theta/l$$

$$\therefore F = \frac{G \cdot 2\pi r \cdot \delta r \cdot r\theta}{l} = \frac{2\pi G\theta r^2 \cdot \delta r}{l}$$

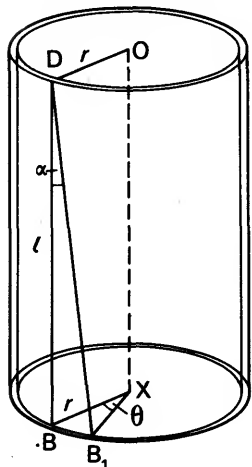


FIG. 6.7
Shear (Torsion) in wire

\therefore moment of F about axis OX of wire $= F \cdot r$

$$= \frac{2\pi G\theta}{l} \cdot r^3 \cdot \delta r.$$

\therefore total moment, or couple torque C ,

$$= \int_0^a \frac{2\pi G\theta}{l} \cdot r^3 dr = \frac{2\pi G\theta}{l} \cdot \frac{a^4}{4}$$

$$\therefore C = \frac{\pi G a^4 \theta}{2l} \quad \dots \dots \dots (i)$$

If the wire is a hollow cylinder of radii a, b respectively, the limits of integration are altered accordingly, and

$$\text{moment of couple} = \int_a^b \frac{2\pi G\theta}{l} \cdot r^3 dr = \frac{\pi G(b^4 - a^4)\theta}{2l}.$$

Determinations of modulus of rigidity. Dynamical method. One method of measuring the modulus of rigidity of a wire E is to clamp it vertically at one end, attach a horizontal disc D of known moment of inertia, I , at the other end, and then time the horizontal torsional oscillations of D . Fig. 6.8 (i). On p. 90, it was shown that the period of oscillation, T , $= 2\pi\sqrt{I/c}$, where c is the opposing couple per unit angle of twist. Thus, with our previous notation, as $\theta = 1$,

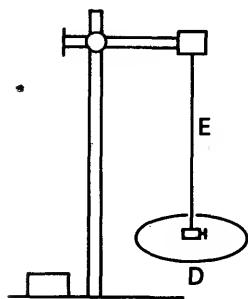
$$c = \frac{\pi G a^4}{2l}.$$

$$\therefore T = 2\pi \sqrt{\frac{2I}{\pi G a^4}}.$$

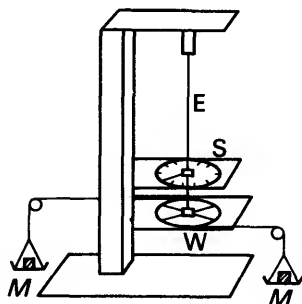
or

$$G = \frac{8\pi I l}{a^4 T^2}$$

Hence G can be evaluated from measurements of l, a, I, T .



(i)



(ii)

FIG. 6.8 Modulus of rigidity measurement

Static method. The modulus of rigidity, G , of the wire E can also be found by measuring the steady deflection θ at the lower end on a scale S graduated in degrees when a couple is applied round a wheel W . Fig. 6.8 (ii). If M is the mass

in each scale-pan, and d is the diameter of W , the moment of the couple on the wire $= Mgd = \pi Ga^4\theta/2l$. The angle θ in radians, and a, l , are known, and hence G can be evaluated.

Poisson's Ratio

When a rubber cord is extended its diameter usually decreases at the same time. *Poisson's ratio*, σ , is the name given to the ratio

$$\frac{\text{lateral contraction/original diameter}}{\text{longitudinal extension/original length}} \quad (1)$$

and is a constant for a given material. If the original length of a rubber strip is 100 cm and it is stretched to 102 cm, the fractional longitudinal extension $= 2/100$. If the original diameter of the cord is 0.5 cm and it decreases to 0.495 cm, the fractional lateral contraction $= 0.005/0.5 = 1/100$. Thus, from the definition of Poisson's ratio,

$$\sigma = \frac{1/100}{2/100} = \frac{1}{2}.$$

When the *volume* of a strip of material remains *constant* while an extension and a lateral contraction takes place, it can easily be shown that Poisson's ratio is 0.5 in this case. Thus suppose that the length of the strip is l and the radius is r .

Then $\text{volume, } V, = \pi r^2 l.$

By differentiating both sides, noting that V is a constant and that we have a product of variables on the right side,

$$\therefore 0 = \pi r^2 \times \delta l + l \times 2\pi r \delta r.$$

$$\therefore r \delta l = -2l \delta r.$$

$$\therefore -\frac{\delta r/r}{\delta l/l} = \frac{1}{2}.$$

But $-\delta r/r$ is the lateral contraction in radius/original radius, and $\delta l/l$ is the longitudinal extension/original length.

$$\therefore \text{Poisson's ratio, } \sigma, = \frac{1}{2}.$$

Experiments show that σ is 0.48 for rubber, 0.29 for steel, 0.27 for iron, and 0.26 for copper. Thus the three metals increase in volume when stretched, whereas rubber remains almost unchanged in volume.

Summary

The three moduli of elasticity are compared in the table below:

	Young's modulus, E	Modulus of Rigidity, G	Bulk modulus, K
1.	Definition : $\frac{\text{tensile stress}}{\text{tensile strain}}$	Definition : $\frac{\text{shear stress}}{\text{shear strain}}$	Definition : $\frac{\text{pressure change}}{-\Delta v/v}$
2.	Relates to change in length ('tensile')	Relates to change in shape ('shear')	Relates to change in volume ('bulk')

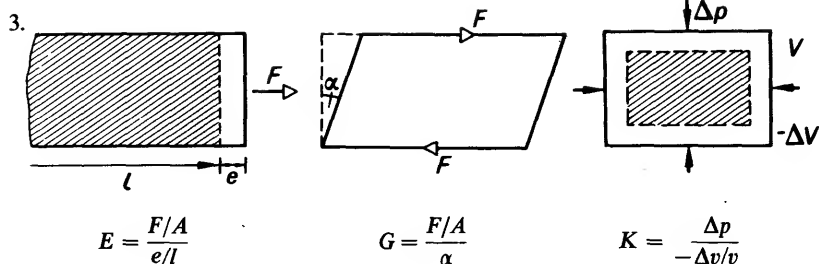


FIG. 6.9

4.	Applies only to solids	Applies to solids and liquids	Applies to all materials. Low value for gases
5.	Used in stretching wires, bending beams, linear expansion and contraction with temperature change	Used in torsion wires, helical springs	Used in velocity of sound formula for all materials. In gas, $K = p$ (isothermal) or γp (adiabatic)

EXERCISES 6

(Assume $g = 9.8 \text{ m s}^{-2}$ unless otherwise stated)

What are the missing words in the statements 1–6?

- When a weight is attached to a suspended long wire, it produces a ... strain.
- The units of Young's modulus are ...
- In measuring Young's modulus, the ... must not be exceeded.
- The energy gained by a wire when stretched = ... \times extension.
- Bulk stress is defined as the ... change.
- When a wire is twisted, a ... strain is produced.

Which of the following answers, A, B, C, D or E, do you consider is the correct one in the statements 9–12?

7. If a metal bar, coefficient of linear expansion α , Young's modulus E , area of cross-section A and length l , is heated through $t^\circ\text{C}$ when clamped at both ends, the force in the bar is calculated from A $EAlt$, B EAt/α , C $EAt\alpha$, D $E^2A\alpha l$, E $A\alpha t l$.

8. When a spiral spring is stretched by a weight attached to it, the strain is A tensile, B shear, C bulk, D elastic, E plastic.

9. The energy in a stretched wire is A $\frac{1}{2}$ load \times extension, B load \times extension, C stress \times extension, D load \times strain, E $\frac{1}{2}$ load \times strain.

10. In an experiment to measure Young's modulus, the wire is thin and long so that A very heavy weights can be attached, B the wire can be suspended from the ceiling, C another identical wire can be arranged parallel to it, D the stress is large and the extension is measurable for laboratory loads, E a micrometer gauge can be used for accurate measurement.

11. Define *tensile stress*, *tensile strain*, *Young's modulus*. What are the units and dimensions of each?

A load of 2 kgf is applied to the ends of a wire 4 m long, and produces an extension of 0.24 mm. If the diameter of the wire is 2 mm, calculate the stress on the wire, its strain, and the value of Young's modulus.

12. What load in kilogrammes must be applied to a steel wire 6 m long and diameter 1.6 mm to produce an extension of 1 mm? (Young's modulus for steel = $2.0 \times 10^{11} \text{ N m}^{-2}$.)

13. Find the extension produced in a copper wire of length 2 m and diameter 3 mm when a load of 3 kgf is applied. (Young's modulus for copper = $1.1 \times 10^{11} \text{ N m}^{-2}$.)

14. What is meant by (i) elastic limit, (ii) Hooke's law, (iii) yield point, (iv) perfectly elastic? Draw sketches of stress *v.* strain to illustrate your answers.

15. 'In an experiment to determine Young's modulus, the strain should not exceed 1 in 1000.' Explain why this limitation is necessary and describe an experiment to determine Young's modulus for the material of a metal wire.

In such an experiment, a brass wire of diameter 0.0950 cm is used. If Young's modulus for brass is $9.86 \times 10^{10} \text{ newton m}^{-2}$, find in kilogram force the greatest permissible load. (L.)

16. Define *stress* and *strain*, and explain why these quantities are useful in studying the elastic behaviour of a material.

State one advantage and one disadvantage in using a long wire rather than a short stout bar when measuring Young's modulus by direct stretching.

Calculate the minimum tension with which platinum wire of diameter 0.1 mm must be mounted between two points in a stout invar frame if the wire is to remain taut when the temperature rises 100K. Platinum has coefficient of linear expansion $9 \times 10^{-6} \text{ K}^{-1}$ and Young's modulus $17 \times 10^{10} \text{ N m}^{-2}$. The thermal expansion of invar may be neglected. (O. & C.)

17. Explain the terms *stress*, *strain*, *modulus of elasticity* and *elastic limit*. Derive an expression in terms of the tensile force and extension for the energy stored in a stretched rubber cord which obeys Hooke's law.

The rubber cord of a catapult has a cross-sectional area 1.0 mm² and a total unstretched length 10.0 cm. It is stretched to 12.0 cm and then released to project a missile of mass 50 g. From energy considerations, or otherwise, calculate the velocity of projection, taking Young's modulus for the rubber as $5.0 \times 10^8 \text{ N m}^{-2}$. State the assumptions made in your calculation.

18. State Hooke's law, and describe in detail how it may be verified experimentally for copper wire. A copper wire, 200 cm long and 1.22 mm diameter, is fixed horizontally to two rigid supports 200 cm long. Find the mass in grams of the load which, when suspended at the mid-point of the wire, produces a sag of 2 cm at that point. Young's modulus for copper = $12.3 \times 10^{10} \text{ N m}^{-2}$. (L.)

19. Distinguish between Young's modulus, the bulk modulus and the shear modulus of a material. Describe a method for measuring Young's modulus. Discuss the probable sources of error and assess the magnitude of the contribution from each.

A piece of copper wire has twice the radius of a piece of steel wire. Young's modulus for steel is twice that for the copper. One end of the copper wire is joined to one end of the steel wire so that both can be subjected to the same longitudinal force. By what fraction of its length will the steel have stretched when the length of the copper has increased by 1%? (O. & C.)

20. In an experiment to measure Young's modulus for steel a wire is suspended vertically and loaded at the free end. In such an experiment, (a) why is the wire long and thin, (b) why is a second steel wire suspended adjacent to the first?

Sketch the graph you would expect to obtain in such an experiment showing the relation between the applied load and the extension of the wire. Show how it is possible to use the graph to determine (a) Young's modulus for the wire, (b) the work done in stretching the wire.

If Young's modulus for steel is $2.00 \times 10^{11} \text{ N m}^{-2}$, calculate the work done in stretching a steel wire 100 cm in length and of cross-sectional area 0.030 cm^2 when a load of 10 kgf is slowly applied without the elastic limit being reached. (N.)

21. Describe the changes which take place when a wire is subjected to a steadily increasing tension. Include in your description a sketch graph of tension against extension for (a) a ductile material such as drawn copper and (b) a brittle one such as cast iron.

Show that the energy stored in a rod of length L when it is extended by a length l is $\frac{1}{2}El^2/L^2$ per unit volume where E is Young's modulus of the material.

A railway track uses long welded steel rails which are prevented from expanding by friction in the clamps. If the cross-sectional area of each rail is 75 cm^2 what is the elastic energy stored per kilometre of track when its temperature is raised by 10°C ? (Coefficient of thermal expansion of steel = $1.2 \times 10^{-5} \text{ K}^{-1}$; Young's modulus for steel = $2 \times 10^{11} \text{ N m}^{-2}$.) (O. & C.)

22. What is meant by saying that a substance is 'elastic'?

A vertical brass rod of circular section is loaded by placing a 5 kg weight on top of it. If its length is 50 cm, its radius of cross-section 1 cm, and the Young's modulus of the material $3.5 \times 10^{10} \text{ N m}^{-2}$, find (a) the contraction of the rod, (b) the energy stored in it. (C.)

23. Give a short account of what happens when a copper wire is stretched under a gradually increasing load. What is meant by *modulus of elasticity*, *elastic limit*, *perfectly elastic*?

When a rubber cord is stretched the change in volume is very small compared with the change in shape. What will be the numerical value of Poisson's ratio for rubber, i.e., the ratio of the fractional decrease in diameter of the stretched cord to its fractional increase in length? (L.)

24. Describe an accurate method of determining Young's modulus for a wire. Detail the precautions necessary to avoid error, and estimate the accuracy attainable.

A steel tyre is heated and slipped on to a wheel of radius 40 cm which it fits exactly at a temperature $t^\circ\text{C}$. What is the maximum value of t if the tyre is not to be stretched beyond its elastic limit when it has cooled to air temperature (17°C)? What will then be the tension in the tyre, assuming it to be 4 cm wide and 3 mm thick? The value of Young's modulus for steel is $1.96 \times 10^{11} \text{ N m}^{-2}$, its coefficient of linear expansion is $1.1 \times 10^{-5} \text{ K}^{-1}$, and its elastic limit occurs for a tension of $2.75 \times 10^8 \text{ N m}^{-2}$. The wheel may be assumed to be at air temperature throughout, and to be incompressible. (O. & C.)

25. State Hooke's law and describe, with the help of a rough graph, the behaviour of a copper wire which hangs vertically and is loaded with a gradually increasing load until it finally breaks. Describe the effect of gradually reducing the load to zero (a) before, (b) after the elastic limit has been reached.

A uniform steel wire of density 7800 kg m^{-3} weighs 16 g and is 250 cm long. It lengthens by 1.2 mm when stretched by a force of 8 kgf. Calculate (a) the value of Young's modulus for the steel, (b) the energy stored in the wire. (N.)

26. Describe an experimental method for the determination of (a) Young's modulus, (b) the elastic limit, of a metal in the form of a thin wire.

A steel rod of mass 97.5 g and of length 50 cm is heated to 200°C and its ends securely clamped. Calculate the tension in the rod when its temperature is reduced to 0°C , explaining how the calculation is made. (Young's modulus for steel $= 2.0 \times 10^{11} \text{ N m}^{-2}$; linear expansivity $= 1.1 \times 10^{-5} \text{ K}^{-1}$; density of steel $= 7800 \text{ kg m}^{-3}$.) (L.)

27. What do you understand by Hooke's law of elasticity? Describe how you would verify it in any particular case.

A wire of radius 0.2 mm is extended by 0.1% of its length when it supports a load of 1 kg; calculate Young's modulus for the material of the wire. (L.)

28. Define Young's modulus of elasticity. Describe an accurate method of determining it. The rubber cord of a catapult is pulled back until its original length has been doubled. Assuming that the cross-section of the cord is 2 mm square, and that Young's modulus for rubber is 10^7 N m^{-2} calculate the tension in the cord. If the two arms of the catapult are 6 cm apart, and the unstretched length of the cord is 8 cm what is the stretching force? (O. & C.)

29. Define *Young's modulus of elasticity* and *coefficient of linear expansion*. State units in which each may be expressed and describe an experimental determination of Young's modulus.

For steel. Young's modulus is $1.8 \times 10^{11} \text{ N m}^{-2}$ and the coefficient of expansion $1.1 \times 10^{-5} \text{ K}^{-1}$. A steel wire 1 mm in diameter is stretched between two supports when its temperature is 200°C . By how much will the force the wire exerts on the supports increase when it cools to 20°C , if they do not yield? Express the answer in terms of the weight of a kilogramme. (L.)

30. Define *elastic limit* and *Young's modulus* and describe how you would find the values for a copper wire.

What stress would cause a wire to increase in length by one-tenth of one per cent if Young's modulus for the wire is $12 \times 10^{10} \text{ N m}^{-2}$? What load in kg wt. would produce this stress if the diameter of the wire is 0.56 mm? (L.)

chapter seven

Solid Friction. Viscosity

SOLID FRICTION

Static Friction

WHEN a person walks along a road, he or she is prevented from slipping by the force of friction at the ground. In the absence of friction, for example on an icy surface, the person's shoe would slip when placed on the ground. The frictional force always *opposes* the motion of the shoe.

The frictional force between the surface of a table and a block of wood A can be investigated by attaching one end of a string to A and the other to a scale-pan S, Fig. 7.1. The string passes over a fixed

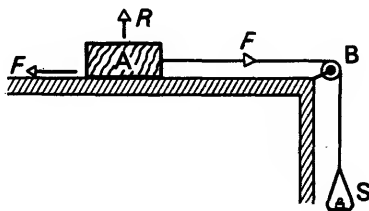


FIG. 7.1 Solid friction

grooved wheel B. When small weights are added to S, the block does not move. The frictional force between the block and table is thus equal to the total weight on S together with the weight of S. When more weights are added, A does not move, showing that the frictional force has increased, but as the weight is increased further, A suddenly begins to slip. The frictional force now present between the surfaces is called the *limiting frictional force*, and we are said to have reached *limiting friction*. The limiting frictional force is the maximum frictional force between the surfaces.

Coefficient of Static Friction

The normal reaction, R , of the table on A is equal to the weight of A. By placing various weights on A to alter the magnitude of R , we can find how the limiting frictional force F varies with R by the experiment just described. The results show that, approximately,

$$\frac{\text{limiting frictional force (F)}}{\text{normal reaction (R)}} = \mu, \text{ a constant,}$$

and μ is known as the *coefficient of static friction* between the two surfaces. The magnitude of μ depends on the nature of the two surfaces; for example, it is about 0.2 to 0.5 for wood on wood, and about 0.2 to 0.6 for wood on metals. Experiment also shows that the limiting frictional force is the same if the block A in Fig. 7.1 is turned on one side so that its surface area of contact with the table decreases, and thus the limiting frictional force is independent of the area of contact when the normal reaction is the same.

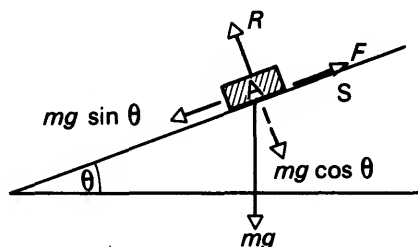


FIG. 7.2 Coefficient by inclined plane

The coefficient of static friction, μ , can also be found by placing the block A on the surface S, and then gently tilting S until A is on the point of slipping down the plane, Fig. 7.2. The static frictional force F is then equal to $mg \sin \theta$, where θ is the angle of inclination of the plane to the horizontal; the normal reaction R is equal to $mg \cos \theta$.

$$\therefore \mu = \frac{F}{R} = \frac{mg \sin \theta}{mg \cos \theta} = \tan \theta,$$

and hence μ can be found by measuring θ .

Kinetic Friction. Coefficient of Kinetic (Dynamic) Friction

When brakes are applied to a bicycle, a frictional force is exerted between the moving wheels and brake blocks. In contrast to the case of static friction, when one of the objects is just on the point of slipping, the frictional force between the moving wheel and brake blocks is called a *kinetic (or dynamic) frictional force*. Kinetic friction thus occurs between two surfaces which have relative motion.

The *coefficient of kinetic (dynamic) friction*, μ' , between two surfaces is defined by the relation

$$\mu' = \frac{F'}{R},$$

where F' is the frictional force when the object moves with a uniform velocity and R is the normal reaction between the surfaces. The coefficient of kinetic friction between a block A and a table can be found by the apparatus shown in Fig. 7.1. Weights are added to the scale-pan, and each time A is given a slight push. At one stage A continues to move with a constant velocity, and the kinetic frictional

force F' is then equal to the total weight in the scale-pan together with the latter's weight. On dividing F' by the weight of A, the coefficient can be calculated. Experiment shows that, when weights are placed on A to vary the normal reaction R , the magnitude of the ratio F'/R is approximately constant. Results also show that the coefficient of kinetic friction between two given surfaces is less than the coefficient of static friction between the same surfaces, and than the coefficient of kinetic friction between two given surfaces is approximately independent of their relative velocity.

Laws of Solid Friction

Experimental results on solid friction are summarised in the *laws of friction*, which state:

(1) The frictional force between two surfaces opposes their relative motion.

(2) The frictional force is independent of the area of contact of the given surfaces when the normal reaction is constant.

(3) The limiting frictional force is proportional to the normal reaction for the case of static friction. The frictional force is proportional to the normal reaction for the case of kinetic (dynamic) friction, and is independent of the relative velocity of the surfaces.

Theory of Solid Friction

The laws of solid friction were known hundreds of years ago, but they have been explained only in comparatively recent years, mainly by F. P. Bowden and collaborators. Sensitive methods, based on electrical conductivity measurements, reveal that the true area of contact between two surfaces is extremely small, perhaps one ten-thousandth of the area actually placed together for steel surfaces. This is explained by photographs which show that some of the atoms of a metal project slightly above the surface, making a number of crests or 'humps'. As Bowden has stated: 'The finest mirror, which is flat to a millionth of a centimetre, would to anyone of atomic size look rather like the South Downs—valley and rolling hills a hundred or more atoms high.' Two metal surfaces thus rest on each others projections when placed one on the other.

Since the area of actual contact is extremely small, the pressures at the points of contact are very high, perhaps 1000 million kgf per m^2 for steel surfaces. The projections merge a little under the high pressure, producing adhesion or 'welding' at the points, and a force which opposes motion is therefore obtained. This explains Law 1 of the laws of solid friction. When one of the objects is turned over, so that a smaller or larger surface is presented to the other object, measurements show that the small area of actual contact remains constant. Thus the frictional force is independent of the area of the surfaces, which explains Law 2. When the load increases the tiny projections are further squeezed

by the enormous pressures until the new area of contact becomes big enough to support the load. The greater the load, the greater is the area of actual contact, and the frictional force is thus approximately proportional to the load, which explains Law 3.

VISCOSITY

If we move through a pool of water we experience a resistance to our motion. This shows that there is a *frictional force* in liquids. We say this is due to the **viscosity** of the liquid. If the frictional force is comparatively low, as in water, the viscosity of the liquid is low; if the frictional force is large, as in glue or glycerine, the viscosity of the liquid is high. We can compare roughly the viscosity of two liquids by filling two measuring cylinders with each of them, and allowing identical small steel ball-bearings to fall through each liquid. The sphere falls more slowly through the liquid of higher viscosity.

As we shall see later, the viscosity of a lubricating oil is one of the factors which decide whether it is suitable for use in an engine. The Ministry of Aircraft Production, for example, listed viscosity values to which lubricating oils for aero-engines must conform. The subject of viscosity has thus considerable practical importance.

Newton's Formula. Coefficient of Viscosity

When water flows slowly and steadily through a pipe, the layer A of the liquid in contact with the pipe is practically stationary, but the central part C of the water is moving relatively fast, Fig. 7.3. At other layers between A and C, such as B, the water has a velocity less than at C, the magnitude of the velocities being represented by the length of the arrowed lines in Fig. 7.3. Now as in the case of two solid surfaces

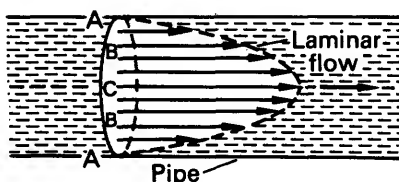


FIG. 7.3 Laminar (uniform) flow through pipe

moving over each other, a frictional force is exerted between two liquid layers when they move over each other. Thus because the velocities of neighbouring layers are different, as shown in Fig. 7.3, a frictional force occurs between the various layers of a liquid when flowing through a pipe.

The basic formula for the frictional force, F , in a liquid was first suggested by NEWTON. He saw that the larger the *area* of the surface of liquid considered, the greater was the frictional force F . He also stated that F was directly proportional to the *velocity gradient* at the part of the liquid considered. This is the case for most common liquids,

called *Newtonian liquids*. If v_1, v_2 are the velocities of C, B respectively in Fig. 7.3, and h is their distance apart, the velocity gradient between the liquids is defined as $(v_1 - v_2)/h$. The velocity gradient can thus be expressed in (m/s)/m, or as 's⁻¹'.

Thus if A is the area of the liquid surface considered, the frictional force F on the surface is given by

$$F \propto A \times \text{velocity gradient},$$

$$\text{or} \quad F = \eta A \times \text{velocity gradient}, \quad (1)$$

where η is a constant of the liquid known as the *coefficient of viscosity*. This expression for the frictional force in a liquid should be contrasted with the case of solid friction, in which the frictional force is independent of the area of contact and of the relative velocity between the solid surfaces concerned (p. 173).

Definition, Units, and Dimensions of Coefficient of Viscosity

The magnitude of η is given by

$$\eta = \frac{F}{A \times \text{velocity gradient}}.$$

The unit of F is a newton, the unit of A is m², and the unit of velocity gradient is 1 m/s per m. Thus η may be defined as *the frictional force per unit area of a liquid when it is in a region of unit velocity gradient*.

The 'unit velocity gradient' = 1 m s⁻¹ change per m. Since the 'm' cancels, the 'unit velocity gradient' = 1 per second. From $\eta = F/(A \times \text{velocity gradient})$, it follows that η may be expressed in units of *newton s m⁻²* (N s m⁻²), or 'dekapoise'.

The coefficient of viscosity of water at 10°C is 1.3×10^{-3} N s m⁻². Since $F = \eta A \times \text{velocity gradient}$, the frictional force on an area of 10 cm² in water at 10°C between two layers of water 0.1 cm apart which move with a relative velocity of 2 cm s⁻¹ is found as follows:

Coefficient of viscosity $\eta = 1.3 \times 10^{-3}$ newton m⁻², $A = 10 \times 10^{-4}$ m², velocity gradient = 2×10^{-2} m s⁻¹ ÷ 0.1×10^{-2} m = $2/0.1$ s⁻¹.

$$\therefore F = 1.3 \times 10^{-3} \times 10 \times 10^{-4} \times 2/0.1 = 2.6 \times 10^{-5} \text{ newton}.$$

Dimensions. The dimensions of a force, F , (= mass \times acceleration = mass \times velocity change/time) are MLT^{-2} . See p. 31. The dimensions of an area, A , are L^2 . The dimensions of velocity gradient

$$= \frac{\text{velocity change}}{\text{distance}} = \frac{\text{L}}{\text{T}} \div \text{L} = \frac{1}{\text{T}}.$$

Now

$$\eta = \frac{F}{A \times \text{velocity gradient}}$$

$$\begin{aligned} \therefore \text{dimensions of } \eta &= \frac{\text{MLT}^{-2}}{\text{L}^2 \times 1/\text{T}} \\ &= \text{ML}^{-1}\text{T}^{-1}. \end{aligned}$$

Thus η may be expressed in units 'kg m⁻¹ s⁻¹'.

Steady Flow of Liquid Through Pipe. Poiseuille's Formula

The steady flow of liquid through a pipe was first investigated thoroughly by POISEUILLE in 1844, who derived an expression for the volume of liquid issuing per second from the pipe. The proof of the formula is given on p. 208, but we can derive most of the formula by the *method of dimensions* (p. 31).

The volume of liquid issuing per second from the pipe depends on (i) the coefficient of viscosity, η , (ii) the radius, a , of the pipe, (iii), the *pressure gradient*, g , set up along the pipe. The pressure gradient = p/l , where p is the pressure difference between the ends of the pipe and l is its length. Thus x, y, z being indices which require to be found, suppose

$$\text{volume per second} = k\eta^x a^y g^z \quad (1)$$

Now the dimensions of volume per second are L^3T^{-1} ; the dimensions of η are $ML^{-1}T^{-1}$, see p. 205; the dimension of a is L ; and the dimensions of g are

$$\frac{[\text{pressure}]}{[\text{length}]}, \text{ or } \frac{[\text{force}]}{[\text{area}][\text{length}]}, \text{ or } \frac{MLT^{-2}}{L^2 \times L}, \text{ which is } ML^{-2}T^{-2}.$$

Thus from (i), equating dimensions on both sides,

$$L^3T^{-1} \equiv (ML^{-1}T^{-1})^x L^y (ML^{-2}T^{-2})^z.$$

Equating the respective indices of M, L, T on both sides, we have

$$x + z = 0,$$

$$-x + y - 2z = 3,$$

$$x + 2z = 1.$$

Solving, we obtain $x = -1, z = 1, y = 4$. Hence, from (1),

$$\text{volume per second} = k \frac{a^4 g}{\eta} = k \frac{pa^4}{l\eta}.$$

We cannot obtain the numerical factor k from the method of dimensions. As shown on p. 209, the factor of $\pi/8$ enters into the formula, which is:

$$\text{Volume per second} = \frac{\pi pa^4}{8\eta l} \quad (2)$$

EXAMPLE

Explain as fully as you can the phenomenon of viscosity, using the viscosity of a gas as the basis of discussion. Show by the method of dimensions how the volume of liquid flowing in unit time along a uniform tube depends on the radius of the tube, the coefficient of viscosity of the liquid, and the pressure gradient along the tube.

The water supply to a certain house consists of a horizontal water main 20 cm in diameter and 5 km long to which is joined a horizontal pipe 15 mm in diameter and 10 m long leading into the house. When water is being drawn by this house

only, what fraction of the total pressure drop along the pipe appears between the ends of the narrow pipe? Assume that the rate of flow of the water is very small. (O. & C.)

Volume per second = $\frac{\pi p a^4}{8\eta l}$, with usual notation.

$$\text{Thus volume per second} = \frac{\pi p_1 \cdot 0.1^4}{8\eta \cdot 5 \times 10^3} = \frac{\pi p_2 \cdot 0.0075^4}{8\eta \cdot 10}$$

where p_1, p_2 are the respective pressures in the two pipes, since the volume per second is the same.

$$\therefore \frac{p_1}{p_2} = \frac{0.0075^4}{0.1} \times \frac{5 \times 10^3}{10} = \frac{1}{63} \text{ (approx.)}$$

$$\therefore p_1 = \frac{1}{64} \times \text{total pressure} = 0.016 \times \text{total pressure}.$$

Turbulent Motion

Poiseuille's formula holds as long as the velocity of each layer of the liquid is parallel to the axis of the pipe and the flow pattern has been developed. As the pressure difference between the ends of the pipe is increased, a critical velocity is reached at some stage, and the motion of the liquid changes from an orderly to a *turbulent* one. Poiseuille's formula does not apply to turbulent motion.

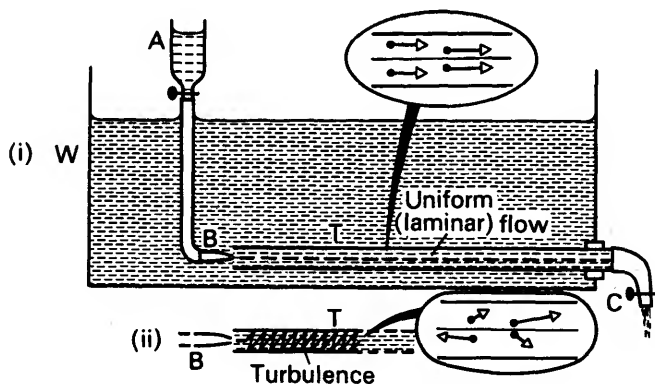


FIG. 7.4 Laminar and turbulent flow

The onset of turbulence was first demonstrated by O. REYNOLDS in 1883, and was shown by placing a horizontal tube T, about 0.5 cm in diameter, at the bottom of a tank W of water, Fig. 7.4 (i). The flow of water along T is controlled by a clip C on rubber tubing connected to T. A drawn-out glass jet B, attached to a reservoir A containing coloured water, is placed at one end of T, and at low velocities of flow a thin coloured stream of water is observed flowing along the middle of T. As the rate of flow of the water along T is increased, a stage is reached

when the colouring in T begins to spread out and fill the whole of the tube, Fig. 7.4 (ii). The critical velocity has now been exceeded, and turbulence has begun.

Fig. 7.4 shows diagrammatically in inset: (i) laminar or uniform flow—here particles of liquid at the same distance from the axis always have equal velocities directed parallel to the axis, (ii) turbulence—here particles at the same distance from the axis have different velocities, and these vary in magnitude and direction with time.

Analogy with Ohm's Law

For orderly flow along a pipe, Poiseuille's formula in equation (2) states:

$$\begin{aligned}\text{Volume per second flowing} &= \frac{\pi p a^4}{8 \eta l}, \\ &= \frac{p \times \pi a^2}{8 \pi \eta \times \frac{l}{\pi a^2}}.\end{aligned}$$

Now $p \times \pi a^2$ = excess pressure \times area of cross-section of liquid = excess force F on liquid, and $l/\pi a^2 = l/A$, where A is the area of cross-section.

$$\therefore \text{volume per second flowing} = \frac{F}{8 \pi \eta \times \frac{l}{A}} \quad (i)$$

The volume of liquid per second is analogous to electric current (I) if we compare the case of electricity flowing along a conductor, and the excess force F is analogous to the potential difference (V) along the conductor. Also, the resistance R of the conductor $= \rho l/A$, where ρ is its resistivity, l is its length, and A is the cross-sectional area. Since, from Ohm's law, $I = V/R$, it follows from (i) that

$8 \pi \eta$ is analogous to ρ , the resistivity;

that is, the coefficient of viscosity η is a measure of the 'resistivity' of a liquid in orderly flow.

Proof of Poiseuille's Formula. Suppose a pipe of radius a has a liquid flowing steadily along it. Consider a cylinder of the liquid of radius r having the same axis as the pipe, where r is less than a . Then the force on this cylinder due to the excess pressure $p = p \times \pi r^2$. We can imagine the cylinder to be made up of cylindrical shells; the force on the cylinder due to viscosity is the algebraic sum of the viscous forces on these shells. The force on one shell is given by $\eta A dv/dr$, where dv/dr is the corresponding velocity gradient and A is the surface area of the shell. And although dv/dr changes as we proceed from the narrowest shell outwards, the forces on the neighbouring shells cancel each other out, by the law of action and reaction, leaving a net force of $\eta A dv/dr$, where dv/dr is the velocity gradient at the surface of the cylinder. The viscous force on the cylinder, and the force on it due to the excess pressure p , are together zero since there is no acceleration of the liquid, i.e., we have orderly or laminar flow.

$$\therefore \eta A \frac{dv}{dr} + \pi r^2 p = 0.$$

$$\therefore \eta \cdot 2\pi r l \frac{dv}{dr} + \pi r^2 p = 0, \text{ since } A = 2\pi r l.$$

$$\therefore \frac{dv}{dr} = -\frac{pr}{2\eta l}$$

$$\therefore v = -\frac{p}{4\eta l} r^2 + c,$$

where c is a constant. Since $v = 0$ when $r = a$, at the surface of the tube, $c = pa^2/4\eta l$.

$$\therefore v = \frac{p}{4\eta l} (a^2 - r^2) \quad \dots \quad (i)$$

Consider a cylindrical shell of the liquid between radii r and $(r + \delta r)$. The liquid in this shell has a velocity v given by the expression in (i), and the volume per second of liquid flowing along this shell $= v \times \text{cross-sectional area of shell}$, since v is the distance travelled in one second, $= v \times 2\pi r \cdot \delta r$.

$$\begin{aligned} \therefore \text{total volume of liquid per second along tube} &= \int_0^a v \cdot 2\pi r \cdot dr \\ &= \int_0^a \frac{p}{4\eta l} (a^2 - r^2) \cdot 2\pi r \cdot dr \\ &= \frac{\pi p a^4}{8\eta l}. \end{aligned}$$

Determination of Viscosity by Poiseuille's Formula

The viscosity of a liquid such as water can be measured by connecting one end of a capillary tube T to a constant pressure apparatus A, which provides a *steady* flow of liquid, Fig. 7.5. By means of a beaker B

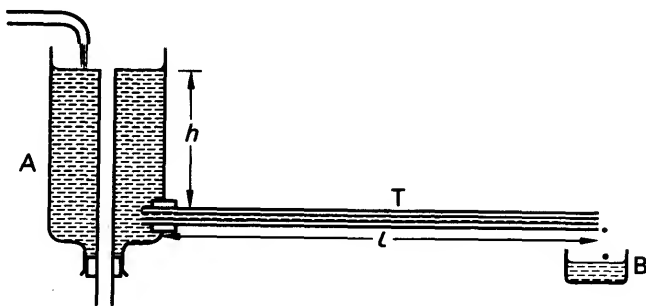


FIG. 7.5 Absolute measurement of viscosity

and a stop-clock, the volume of water per second flowing through the tube can be measured. The pressure difference between the ends of T is $h\rho g$, where h is the pressure head, ρ is the density of the liquid, and g is 9.8 m s^{-2} .

$$\therefore \text{volume per second} = \frac{\pi p a^4}{8\eta l} = \frac{\pi h \rho g a^4}{8\eta l},$$

where l is the length of T and a is its radius. The radius of the tube can be measured by means of a mercury thread or by a microscope. The coefficient of viscosity η can then be calculated, since all the other quantities in the above equation are known.

Comparison of Viscosities. Ostwald Viscometer

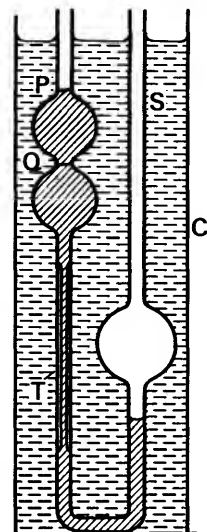


FIG. 7.6
Ostwald viscometer

An Ostwald viscometer, which contains a vertical capillary tube T, is widely used for comparing the viscosities of two liquids, Fig. 7.6. The liquid is introduced at S, drawn by suction above P, and the time t_1 taken for the liquid level to fall between the fixed marks P, Q is observed. The experiment is then repeated with the *same volume* of a second liquid, and the time t_2 for the liquid level to fall from P to Q is noted.

Suppose the liquids have respective densities ρ_1, ρ_2 . Then, since the average head h of liquid forcing it through T is the same in each case, the pressure excess between the ends of T = $h\rho_1g, h\rho_2g$ respectively. If the volume between the marks P, Q is V , then, from Poiseuille's formula, we have

$$\frac{V}{t_1} = \frac{\pi(h\rho_1g)a^4}{8\eta_1l} \quad (i),$$

where a is the radius of T, η_1 is the coefficient of viscosity of the liquid, and l is the length of T. Similarly, for the second liquid,

$$\frac{V}{t_2} = \frac{\pi(h\rho_2g)a^4}{8\eta_2l} \quad (ii)$$

Dividing (ii) by (i),

$$\begin{aligned} \therefore \frac{t_1}{t_2} &= \frac{\eta_1\rho}{\eta_2\rho_1} \\ \therefore \frac{\eta_1}{\eta_2} &= \frac{t_1}{t_2} \cdot \frac{\rho_1}{\rho_2} \quad (iii) \end{aligned}$$

Thus knowing t_1, t_2 and the densities ρ_1, ρ_2 , the coefficients of viscosity can be compared. Further, if a pure liquid of a known viscosity is used, the viscometer can be used to measure the coefficient of viscosity of a liquid. Since the viscosity varies with temperature, the viscometer should be used in a cylinder C and surrounded by water at a constant temperature, Fig. 7.6. The arrangement can then also be used to investigate the variation of viscosity with temperature. In very accurate work a small correction is required in equation (iii). BARR, an authority on viscosity, estimates that nearly 90% of petroleum oil is tested by an Ostwald viscometer.

Experiment shows that the viscosity coefficient of a liquid diminishes as its temperature rises. Thus for water, η at 15°C is $1.1 \times 10^{-3} \text{ N s m}^{-2}$, at 30°C it is $0.8 \times 10^{-3} \text{ N s m}^{-2}$ and at 50°C it is $0.6 \times 10^{-3} \text{ N s m}^{-2}$. Lubricating oils for motor engines which have the same coefficient of viscosity in summer and winter are known as ‘viscostatic’ oils.

Stokes' Law, Terminal Velocity

When a small object, such as a steel ball-bearing, is dropped into a viscous liquid like glycerine it accelerates at first, but its velocity soon reaches a steady value known as the *terminal velocity*. In this case the viscous force acting upwards, and the upthrust due to the liquid on the object, are together equal to its weight acting downwards, so that the resultant force on the object is zero. An object dropped from an aeroplane at first increases its speed v , but soon reaches its terminal speed. Fig. 7.7 shows that variation of v with time as the terminal velocity v_0 is reached.

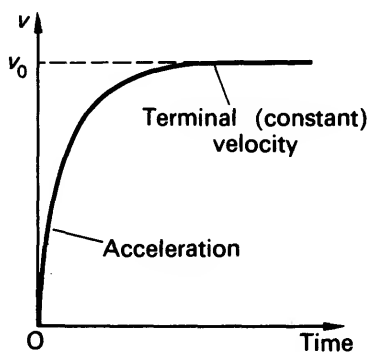


FIG. 7.7 Motion of falling sphere

Suppose a sphere of radius a is dropped into a viscous liquid of coefficient of viscosity η , and its velocity at an instant is v . The frictional force, F , can be partly found by the method of dimensions. Thus suppose $F = ka^x\eta^y v^z$, where k is a constant. The dimensions of F are MLT^{-2} ; the dimension of a is L ; the dimensions of η are $\text{ML}^{-1}\text{T}^{-1}$; and the dimensions of v are LT^{-1} .

$$\therefore \text{MLT}^{-2} \equiv \text{L}^x \times (\text{ML}^{-1}\text{T}^{-1})^y \times (\text{LT}^{-1})^z.$$

Equating indices of M, L, T on both sides,

$$\therefore y = 1,$$

$$x - y + z = 1,$$

$$-y-z = -2.$$

Hence $z = 1$, $x = 1$, $y = 1$. Consequently $F = k\eta av$. In 1850 STOKES showed mathematically that the constant k was 6π , and he arrived at the formula

$$F = 6\pi a\eta v \quad . \quad . \quad . \quad (1)$$

Comparison of Viscosities of Viscous Liquids

Stokes' formula can be used to compare the coefficients of viscosity of very viscous liquids such as glycerine or treacle. A tall glass vessel *G* is filled with the liquid, and a small ball-bearing *P* is dropped gently into the liquid so that it falls along the axis of *G*, Fig. 7.8. Towards the middle of the liquid *P* reaches its terminal velocity v_0 , which is measured by timing its fall through a distance *AB* or *BC*.

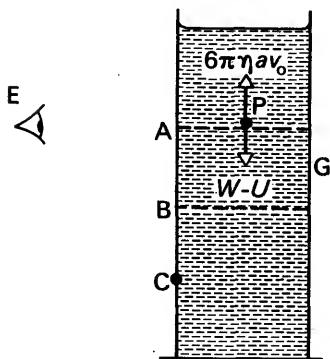


FIG. 7.8 Stokes' law

The upthrust, *U*, on *P* due to the liquid $= 4\pi a^3 \sigma g/3$, where *a* is the radius of *P* and σ is the density of the liquid. The weight, *W*, of *P* is $4\pi a^3 \rho g/3$, where ρ is density of the bearing's material. The net downward force is

thus $4\pi a^3 g(\rho - \sigma)/3$. When the opposing frictional force grows to this magnitude, the resultant force on the bearing is zero. Thus for the terminal velocity v_0 , we have

$$6\pi\eta av_0 = \frac{4}{3}\pi a^3 g(\rho - \sigma),$$

$$\therefore \eta = \frac{2ga^2(\rho - \sigma)}{9v_0} \quad \dots \dots \dots (i)$$

When the experiment is repeated with a liquid of coefficient of viscosity η_1 and density σ_1 , using the same ball-bearing, then

$$\eta_1 = \frac{2ga^2(\rho - \sigma_1)}{9v_1} \quad \dots \dots \dots (ii)$$

where v_1 is the new terminal velocity. Dividing (i) by (ii),

$$\therefore \frac{\eta}{\eta_1} = \frac{v_1(\rho - \sigma)}{v_0(\rho - \sigma_1)} \quad \dots \dots \dots (iii)$$

Thus knowing v_1 , v , ρ , σ_1 , σ , the coefficients of viscosity can be compared. In very accurate work a correction to (iii) is required for the effect of the walls of the vessel containing the liquid.

Molecular theory of viscosity

Viscous forces are detected in gases as well as in liquids. Thus if a disc is spun round in a gas close to a suspended stationary disc, the latter rotates in the same direction. The gas hence transmits frictional forces. The flow of gas through pipes, particularly in long pipes as in transmission of natural gas from the North Sea area, is affected by the viscosity of the gas.

The viscosity of *gases* is explained by the transfer of momentum which

takes place between neighbouring layers of the gas as it flows in a particular direction. Fast-moving molecules in a layer X cross with their own velocity to a layer Y say where molecules are moving with a slower velocity. Fig. 7.9. Molecules in Y likewise move to X. The net

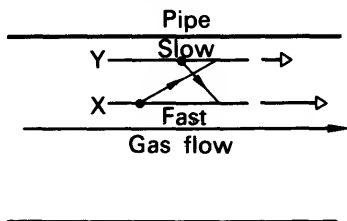


FIG. 7.9 Viscosity of gas—momentum effect

effect is an increase in momentum in Y and a corresponding decrease in X, although on the average the total number of molecules in the two layers is unchanged. Thus the layer Y speeds up and the layer X slows down, that is, a *force* acts on the layers of the gas while they move. This is the viscous force. We consider the movement of molecules in more detail shortly.

Although there is transfer of momentum as in the gas, the viscosity of a *liquid* is mainly due to the molecular attraction between molecules in neighbouring layers. Energy is needed to drag one layer over the other against the force of attraction. Thus a shear stress is required to make the liquid move in laminar flow.

EXERCISE 7

What are the missing words in the statements 1–6?

1. The coefficient of dynamic (kinetic) friction is the ratio . . .
2. The coefficient of friction between two given surfaces is . . . of the area in contact.
3. In orderly or laminar flow of liquids in a pipe, the volume per second flowing past any section is given by the formula . . .
4. The dimensions of coefficient of viscosity are . . .
5. When a small sphere of radius a falls through a liquid with a constant velocity v , the frictional force is given by the formula . . .
6. In comparing the viscosities of water and alcohol by an Ostwald viscometer, the same liquid . . . must used.

Which of the following answers, A, B, C, D or E, do you consider is the correct one in the statements 7–10?

7. In orderly or laminar flow of a liquid through a pipe, *A* tensile forces act on the layers and the volume per second V is proportional to the pressure at one end, *B* shear forces act on the layers and V is proportional to the pressure at one end, *C* shear forces act on the layers and V is proportional to the pressure difference between the ends, *D* bulk forces act throughout the liquid, *E* V is directly proportional to a^4 and to the coefficient of viscosity.

8. When a small steel sphere is dropped gently down the axis of a wide jar of glycerine, the sphere *A* travels with constant velocity throughout its motion, *B* accelerates at first and then reaches a constant velocity, *C* decelerates at first and then reaches a constant velocity, *D* accelerates throughout its motion, *E* slowly comes to rest.

9. When a gas flows steadily along a pipe, the viscous forces in it are due to *A* transfer of energy from one layer to another, *B* the uniform speed of the molecules, *C* the varying density along the pipe, *D* the transfer of momentum from one layer to another, *E* the varying pressure at a given section of the pipe.

10. A pipe *P* has twice the diameter of a pipe *Q*, and *P* has a liquid *X* flowing along it which has twice the viscosity of a liquid *Y* flowing through *Q*. If the flow is orderly or laminar in each, and the volume per second in *P* and *Q* is the same, the pressure difference at the ends of *P* compared to that of *Q* is *A* 1:8, *B* 1:4, *C* 8:1, *D* 4:1, *E* 1:1.

Solid Friction

11. State the laws of solid friction. Describe an experiment to determine the coefficient of dynamic (or sliding) friction between two surfaces.

A horizontal circular turntable rotates about its centre at the uniform rate of 120 revolutions per minute. Find the greatest distance from the centre at which a small body will remain stationary relative to the turntable, if the coefficient of static friction between the turntable and the body is 0.80. (*L*.)

12. State (*a*) the laws of solid friction, (*b*) the triangle law for forces in equilibrium. Describe an experiment to determine the coefficient of sliding (dynamic) friction between two wooden surfaces.

A block of wood of mass 150 g rests on an inclined plane. If the coefficient of static friction between the surfaces in contact is 0.30, find (*a*) the greatest angle to which the plane may be tilted without the block slipping, (*b*) the force parallel to the plane necessary to prevent slipping when the angle of the plane with the horizontal is 30° , showing that this direction of the force is the one for which the force required to prevent slipping is a minimum. (*L*.)

13. Distinguish between *static* and *sliding* (kinetic) friction and define the *coefficient of sliding friction*.

How would you investigate the laws of sliding friction between wood and iron?

An iron block, of mass 10 kg, rests on a wooden plane inclined at 30° to the horizontal. It is found that the least force parallel to the plane which causes the block to slide *up* the plane is 10 kgf. Calculate the coefficient of sliding friction between wood and iron. (*N*.)

14. Give an account of the factors which determine the force of friction (i) between solids, (ii) in liquids.

A block of mass 12 kg is drawn along a horizontal surface by a steadily applied force of 4 kg weight acting in the direction of motion. Find the kinetic energy acquired by the block at the end of 10 seconds and compare it with the total work done on the block in the same time. (Coefficient of friction = 0.28.) (*L*.)

15. State the laws of solid friction.

Describe experiments to verify these laws, and to determine the coefficient of static friction, for two wooden surfaces.

A small coin is placed on a gramophone turntable at a distance of 7.0 cm from the axis of rotation. When the rate of rotation is gradually increased from zero

the coin begins to slide outwards when the rate reaches 60 revolutions per minute. Calculate the rate of rotation for which sliding would commence if (a) the coin were placed 12.0 cm from the axis, (b) the coin were placed in the original position with another similar coin stuck on top of it. (L)

16. Define *coefficient of sliding friction*, *coefficient of viscosity*. Contrast the laws of solid friction with those which govern the flow of liquids through tubes.

Sketch the apparatus you would employ to determine the coefficient of sliding friction between a wood block and a board and show how you would deduce the coefficient from a suitable graph. (L.)

Viscosity

17. Define *coefficient of viscosity* of a fluid.

When the flow is orderly the volume V of liquid which flows in time t through a tube of radius r and length l when a pressure difference p is maintained between its ends is given by the equation $\frac{V}{t} = \frac{\pi pr^4}{8l\eta}$ where η is the coefficient of viscosity

of the liquid. Describe an experiment based on this equation *either* (a) to determine the value of η for a liquid, *or* (b) to compare the values of η for two liquids, pointing out the precautions which must be taken in the experiment chosen to obtain an accurate result.

Water flows steadily through a horizontal tube which consists of two parts joined end to end; one part is 21 cm long and has a diameter of 0.225 cm and the other is 7.0 cm long and has a diameter of 0.075 cm. If the pressure difference between the ends of the tube is 14 cm of water find the pressure difference between the ends of each part. (L.)

18. The dimensions of *energy*, and also those of *moment of a force* are found to be 1 in mass, 2 in length and -2 in time. Explain and justify this statement.

(a) A sphere of radius a moving through a fluid of density ρ with high velocity V experiences a retarding force F given by $F = k \cdot a^x \cdot \rho^y \cdot V^z$, where k is a non-dimensional coefficient. Use the method of dimensions to find the values of x , y and z .

(b) A sphere of radius 2 cm and mass 100 g, falling vertically through air of density 1.2 kg m^{-3} , at a place where the acceleration due to gravity is 9.81 m s^{-2} , attains a steady velocity of 30 m s^{-1} . Explain why a constant velocity is reached and use the data to find the value of k in this case. (O. & C.)

19. Mass, length and time are *fundamental* units, whereas acceleration, force and energy are *derived* units. Explain the distinction between these two types of unit. Define each of the three derived units and apply your definition in each case to deduce its dimensions.

An incompressible fluid of viscosity η flows along a straight tube of length l and uniform circular cross-section of radius r . Provided the pressure difference p between the ends of the tube is not too great the velocity u of fluid flow along the axis of the tube is found to be directly proportional to p . Apply the method of dimensions to deduce this result assuming u depends only on r , l , η and p .

How may the viscosity of an ideal gas be accounted for by elementary molecular theory? (O. & C.)

20. Define *coefficient of viscosity*. Describe an experiment to compare the coefficients of viscosity of water and benzene at room temperature.

A small metal sphere is released from rest in a tall wide vessel of liquid. Discuss the forces acting on the sphere (a) at the moment of release, (b) soon after release, (c) after the terminal velocity has been attained.

Castor oil at 20°C has a coefficient of viscosity 2.42 N s m^{-2} and a density 940 kg m^{-3} . Calculate the terminal velocity of a steel ball of radius 2.0 mm falling under gravity in the oil, taking the density of steel as 7800 kg m^{-3} . (L.)

21. Define *coefficient of viscosity*. What are its dimensions?

By the method of dimensions, deduce how the rate of flow of a viscous liquid through a narrow tube depends upon the viscosity, the radius of the tube, and the pressure difference per unit length. Explain how you would use your results to compare the coefficients of viscosity of alcohol and water. (C.)

22. Define *coefficient of viscosity*. For orderly flow of a given liquid through a capillary tube of length l , radius r , the volume of liquid issuing per second is proportional to pr^4/l where p is the pressure difference between the ends of the tube. How would you verify this relation experimentally for water at room temperature? How would you detect the onset of turbulence? (N.)

23. The viscous force acting on a small sphere of radius a moving slowly through a liquid of viscosity η with velocity v is given by the expression $6\pi\eta av$. Sketch the general shape of the velocity-time graph for a particle falling from rest through a viscous fluid, and explain the form of the graph. List the observations you would make to determine the coefficient of viscosity of the fluid from the motion of the particle.

Some particles of sand are sprinkled on to the surface of the water in a beaker filled to a depth of 10 cm . Estimate the least time for which grains of diameter 0.10 mm remain in suspension in the water, stating any assumptions made.

[Viscosity of water $= 1.1 \times 10^{-3} \text{ N s m}^{-2}$; density of sand $= 2200 \text{ kg m}^{-3}$.] (C.)

24. Define *coefficient of friction* and *coefficient of viscosity*.

Describe how you would (a) measure the coefficient of sliding friction between iron and wood, and (b) compare the viscosities of water and paraffin oil. (L.)

PART TWO

Heat

chapter eight

Introduction

Temperature

WE are interested in heat because it is the commonest form of energy, and because changes of temperature have great effects on our personal comfort, and on the properties of substances, such as water, which we use every day. *Temperature* is a scientific quantity which corresponds to primary sensations—hotness and coldness. These sensations are not reliable enough for scientific work, because they depend on

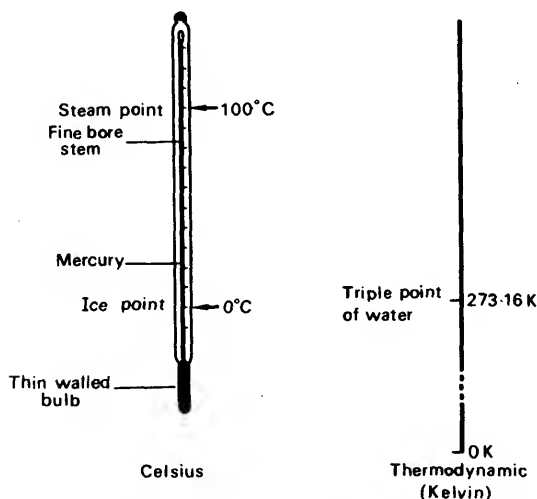


FIG. 8.1. Mercury-in-glass thermometer (left); °C and K scales.

contrast—the air in a thick-walled barn or church feels cool on a summer's day, but warm on a winter's day, although a thermometer may show that it has a lower temperature in the winter. A thermometer, such as the familiar mercury-in-glass instrument (Fig. 8.1), is a device whose readings depend on hotness or coldness, and which we choose to consider more reliable than our senses. We are justified in considering it more reliable because different thermometers of the same type agree with one another better than different people do.

The temperature of a body, then, is its degree of hotness, as measured on a thermometer. The thermometer was invented in Italy about 1630: it consisted of an open-ended tube, with a bulb full of water at its lower end. The water rose in the tube when the bulb was warmed, and fell when it was cooled.

As a liquid for use in thermometers, water soon gave way to linseed oil or alcohol, and by about 1660 thermometer-makers had begun to seal the top of the tube. Early thermometers had no definite scale, like that of a modern thermometer; some of them were used for showing the temperatures of greenhouses, and were mounted on wooden backboards, which were carved with grapes and peaches for example, to indicate the correct temperatures for growing the different fruits. The thermometer as we know it to-day, containing pure mercury and graduated according to a universal scale, was developed by Fahrenheit in 1724.

Temperature Scales. Celsius

When a mercury thermometer is to be graduated, it is placed first in melting ice, and then in steam from boiling water (Fig. 8.2). The temperature of the steam depends on the atmospheric pressure, as we

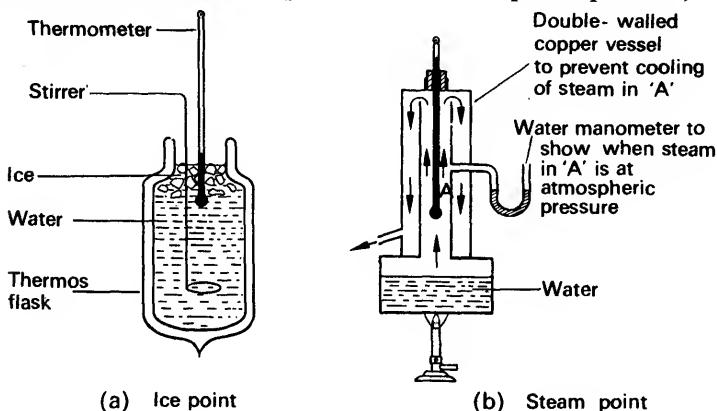


FIG. 8.2. Determination of fixed points.

shall see in Chapter 12; for calibrating thermometers, an atmospheric pressure of 76 cm mercury is chosen. In both steam and ice, when the level of the mercury has become steady, it is marked on the glass: the level in ice is called the *lower fixed point*, or ice-point, and the level in steam is called the *upper fixed point*, or steam-point. The distance between the fixed points is called the *fundamental interval* of the thermometer. For scientific work, the fundamental interval is divided into 100 equal parts (Fig. 8.1). This division was first proposed by Celsius in 1742, and the graduations are called degrees Centigrade or, in modern nomenclature, degrees Celsius ($^{\circ}\text{C}$); the ice-point is 0°C and the steam-point 100°C .

Thermodynamic Scale

The *thermodynamic scale* of temperature is adopted as the SI temperature scale. On this scale, the *kelvin* is the unit of temperature. It is defined as $1/273.16$ of the thermodynamic temperature of the triple point of water (p. 319).

The symbol for temperature is 'K' without a degree sign. Thus the triple point of water, T_{tr} , is 273.16 K exactly. The absolute zero, 0 K,

is -273.15°C . To a good approximation, $0^{\circ}\text{C} = 273\text{ K}$ and $100^{\circ}\text{C} = 373\text{ K}$. (Fig. 8.1.)

The temperature change or interval of one degree Celsius, 1 degC or 1°C , is exactly the same as the temperature interval of one degree on the thermodynamic scale. On this account the interval 'degC or $^{\circ}\text{C}$ ' is written 'K' in SI units. Similarly, 'per degC or per $^{\circ}\text{C}$ ' is written ' K^{-1} ' in SI units. For example, the linear expansivity (formerly, linear coefficient of expansion) of steel is written ' $12 \times 10^{-6}\text{ K}^{-1}$ ' in SI units, in place of ' 12×10^{-6} per degC'. The use of ' K^{-1} ' occurs frequently in units throughout the subject and should be noted by the reader.

Types of Thermometer

The *mercury-in-glass* thermometer depends on the change in volume of the mercury with hotness; it is cheap and simple, but is not reliable enough for accurate work (Chapter 14). Other types of thermometer depend on the change, with hotness, of the pressure of a gas at constant volume or the electrical resistance of a metal (Fig. 8.3). Another type of

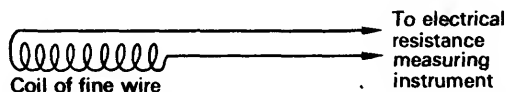


FIG. 8.3. A resistance thermometer; the wire is usually of platinum.

thermometer depends on the electromotive force change with temperature of two metals joined together. Fig. 8.4 (a) shows two wires, one of copper and one of iron, soldered together at A. The ends of the wires are joined to a galvanometer G. When the junction A is heated, a current flows which deflects the galvanometer. The current usually increases with the temperature difference between the hot and cold ends of the wires. For temperature measurement two junctions are used, as in Fig. 8.4 (b); the second one, called the cold junction, is maintained at 0°C by ice-water.

Each of these quantities—e.m.f., resistance, pressure—gives its own temperature scale, and the different scales agree only at the fixed points, where their readings are *defined* as 0°C and 100°C . (When we speak of

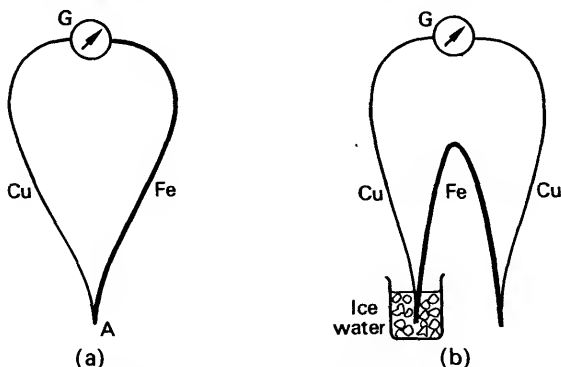


FIG. 8.4. Thermojunctions or thermocouples.

a temperature *scale*, we refer to the quantity used to define it; the difference between °C and K is only a difference in the *graduation* of a given scale.) If, for example, a given platinum wire has resistances R_0 and R_{100} at the ice- and steam-point respectively, then its fundamental interval is $R_{100} - R_0$. And if it has a resistance R at an unknown temperature, the value of that temperature, t_p , on the platinum resistance Celsius scale, is given by

$$t_p = \frac{R - R_0}{R_{100} - R_0} \times 100(^{\circ}\text{C}).$$

The platinum-resistance scale differs appreciably from the mercury-in-glass scale, as the following table shows:

Mercury-in-glass	0	50	100	200	300	°C
Platinum-resistance	0	50.25	100	197	291	°C

We shall discuss temperature scales again later (p. 366); here we wish only to point out that they differ from one another, that no one of them is any more 'true' than any other, and that our choice of which to adopt is arbitrary, though it may be decided by convenience.

Effects of Temperature

Most bodies, when they are made hotter, become larger; their increase in size is called thermal expansion. Thermal expansion may be useful, as in a thermometer, or it may be a nuisance, as in bridges and railway lines. If the thermal expansion of a solid or liquid is resisted, great forces are set up: that is why gaps are left between railway lines, and why beer-bottles are never filled quite full. If the thermal expansion of a gas is resisted, however, the forces set up are not so great; the pressure of the gas increases, but not catastrophically. The increase of pressure is, in fact, made use of in most forms of engine; it is also made use of in accurate thermometry.

Besides causing a change in size or pressure, a change of temperature

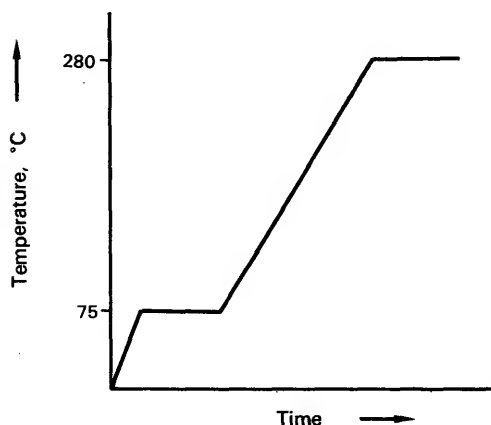


FIG. 8.5. Warming curve of lead acetate.

may cause a change of state—from solid to liquid, liquid to gas, or vice versa. If we heat some crystals of lead acetate in a crucible, and measure their temperature, with a thermometer reading to 300°C, we find that the crystals warm steadily up to 75°C and then start to melt. Their temperature does not rise further until they have all melted (Fig. 8.5). After it has melted, the lead acetate warms up to 280°C, and

then keeps a constant temperature until it has all boiled away. We call 280°C the boiling-point of lead acetate; likewise we call 75°C the melting- (or freezing-) point of lead acetate.

HEAT AND ENERGY

Heat and Temperature

If we run hot water into a lukewarm bath, we make it hotter; if we run cold water in, we make it cooler. The hot water, we say, gives heat to the cooler bath-water; but the cold water takes heat from the warmer bath-water. The quantity of heat which we can get from hot water depends on both the mass of water and on its temperature: a bucket-full at 80°C will warm the bath more than a cup-full at 100°C . Roughly speaking, temperature is analogous to electrical potential, and heat is analogous to quantity of electricity. We can perceive temperature changes, and whenever the temperature of a body rises, that body has gained heat. The converse is not always true; when a body is melting or boiling, it is absorbing heat from the flame beneath it, but its temperature is not rising.

Latent Heat and Specific Heat

The heat which a body absorbs, in melting or boiling, it gives out again in freezing or condensing; such heat is called *latent*, or hidden, *heat*, because it does not show itself by a change in temperature. When a body absorbs heat without changing its state, its temperature rises, and the heat absorbed was first called 'sensible heat'.

The term 'latent heat' was used by Black (1728–99); he and a Swede, Wilcke, discovered latent heats independently at about the same date—Black by hanging a bucket of ice in a warm room, Wilcke by pouring boiling water on to snow.

Also independently, Black and Wilcke studied what we now call *specific heats*; the name is due to Wilcke. In his experiments Wilcke dropped various hot bodies into cold water, and measured the temperature rises which they caused. In this way he showed that a given mass of glass, for example, gave out only one-fifth as much heat as an equal mass of water, in cooling through the same temperature range. He therefore said that the specific heat of glass was 0.2.

In the seventeenth and eighteenth centuries the nature of heat was disputed; some thought of heat as the motion of the particles of a body, others thought of it as a fluid, filling the body's pores. Measurements of heat were all relative, and no unit of the quantity of heat was defined. In the nineteenth century, however, the increasing technical importance of heat made a unit of it essential. The units of heat chosen were:

- (i) the *calorie* (cal): this is the amount of heat required to warm 1 gramme (g) of water through 1°C (see also p. 194);
- (ii) the *British Thermal Unit* (Btu): this is the amount of heat required to warm 1 lb of water through 1°F .

Heat and Energy

Steam-engines became common in the early part of the eighteenth century; but they were not thought of as heat-engines until the latter part of that century. Consequently the early engines were wasteful of fuel, squandering useful heat in warming and cooling the cylinder at every stroke of the piston. Watt reduced this waste of heat by his invention of the separate condenser in 1769. Trevithick, about 1800, devised an engine which was driven by steam which entered the cylinder at a pressure above atmospheric, and therefore at a temperature above 100°C (p. 304). In this engine, the steam came out of the exhaust at a temperature no higher than in earlier engines, so that a greater fraction of the heat which it carried from the boiler was used in the engine.

The idea of heat as a form of energy was developed particularly by Benjamin Thompson (1753–1814); he was an American who, after adventures in Europe, became a Count of the Holy Roman Empire, and war minister of Bavaria. He is now generally known as Count Rumford. While supervising his arsenal, he noticed the great amount of heat which was liberated in the boring of cannon. The idea common at the time was that this heat was a fluid, pressed out of the chips of metal as they were bored out of the barrel. To measure the heat produced, Rumford used a blunt borer, and surrounded it and the end of the cannon with a wooden box, filled with water (Fig. 8.6). From the weight of water, and the rate at which its temperature rose, he concluded that the boring operation liberated heat at the same rate as 'nine wax candles, burning with a clear flame'. He showed that the amount of heat liberated was in no way connected with the mass of metal bored away, and concluded that it depended only on the work done against friction. It followed that **heat was a form of energy**.

Rumford published the results of his experiments in 1798. No similar experiments were made until 1840, when Joule began his study of heat and other forms of energy. Joule measured the work done, and the heat produced, when water was churned, in an apparatus which we shall describe on p. 197. He also measured the work done and heat produced when oil was churned, when air was compressed, when water was forced through fine tubes, and when cast iron bevel wheels were rotated one against the other. Always, within the limits of experimental error, he found that the heat liberated was proportional to the mechanical work done, and that the ratio of the two was the same in all types of experiment. His last experiments, made in 1878, showed that about 772 ft-lbf of work were equivalent to one British thermal unit of heat. This ratio Joule called the *mechanical equivalent of heat*. The metric unit of work or energy is the *joule*, J. Since experiment shows that heat is a form of energy, *the joule is now the scientific unit of heat*. 'Heat per second' is expressed in 'joules per second' or *watts*, W.

Today, from definition, about $4.2 \text{ J} = 1 \text{ calorie}$ (more accurately, $4.187 \text{ J} = 1 \text{ calorie}$). Calories are units still used by chemists, for example. Values in calories met can be converted to joules approximately by multiplying them by 4.2. See p. 199.

In other experiments, Joule measured the heat liberated by an electric

current in flowing through a resistance; at the same time he measured the work done in driving the dynamo which generated the current. He obtained about the same ratio for work done to heat liberated as in his direct experiments. This work linked the ideas of heat, mechanical, and electrical energy. He also showed that the heat produced by a current is related to the chemical energy used up.

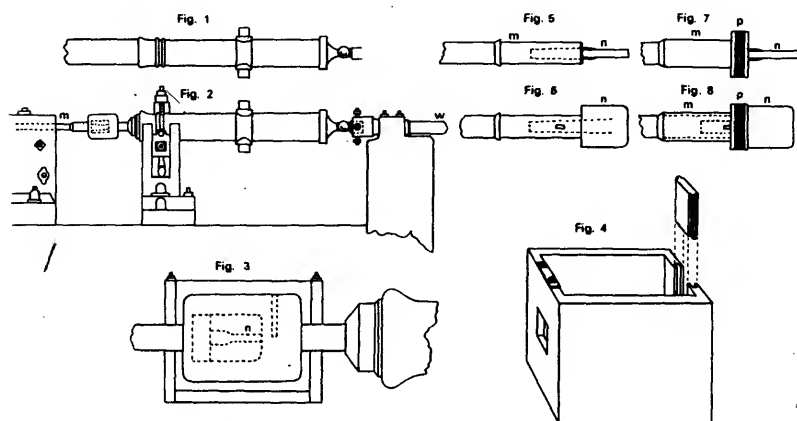


FIG. 8.6. Rumford's apparatus for converting work into heat.

Fig. 1 shows the cannon. Fig. 2 shows the complete apparatus; *w* is connected to machinery driven round by horses, *m* is joined to a blunt borer *n* in a cylinder shown enlarged in Fig. 3 and Fig. 4. Figs. 5, 6, 7, 8 show further details of *m* and *n*.

The Conservation of Energy

As a result of all his experiments, Joule developed the idea that energy in any one form could be converted into any other. There might be a loss of useful energy in the process—for example, some of the heat from the furnace of a steam-engine is lost up the chimney, and some more down the exhaust—but no energy is destroyed. The work done by the engine added to the heat lost as described and the heat developed as friction, it is equal to the heat provided by the fuel burnt. The idea underlying this statement is called the *Principle of the Conservation of Energy*. It implies that, if we start with a given amount of energy in any one form, we can convert it in turn into all other forms; we may not always be able to convert it completely, but if we keep an accurate balance-sheet we shall find that the total amount of energy, expressed in any one form—say heat or work—is always the same, and is equal to the original amount.

The conservation of energy applies to living organisms—plants and animals—as well as to inanimate systems. For example, we may put a man or a mouse into a box or a room, give him a treadmill to work, and feed him. His food is his fuel; if we burn a sample of it, we can measure its chemical energy, in heat units. And if we now add up the

heat value of the work which the man does, and the heat which his body gives off, we find that their total is equal to the chemical energy of the food which the man eats. Because food is the source of man's energy, food values are commonly expressed in *kilocalories*, which is the heat required to warm 1 kilogramme of water through 1 deg C. A man needs about 3000 kilocalories per day.

Muscles are unique in their capacity to turn chemical energy directly into mechanical energy. When a muscle is stimulated, complex phosphates in its tissues break down; in doing so, they cause the muscle fibres to swell and shorten. Thus, via the bones and joints, the muscle does external work. When the muscle is recovering after contraction, the phosphates are built up again by a series of reactions, involving the oxidation of sugars. The sugars and oxygen are brought to the muscle in the arterial blood; the waste products of the reactions, water and carbon dioxide, are carried away in the venous blood.¹ Recently physiologists have found evidence that muscles may also convert mechanical energy into chemical.² For example, when we walk downstairs, gravity does work on our leg-muscles; some of this appears as heat, but some, it now seems, is used in reversing the chemical actions of muscle activity.

All the energy by which we live comes from the sun. The sun's ultra-violet rays are absorbed in the green matter of plants, and make them grow; the animals eat the plants, and we eat them—we are all vegetarians at one remove. The plants and trees of an earlier age decayed, were buried, and turned into coal. Even water-power comes

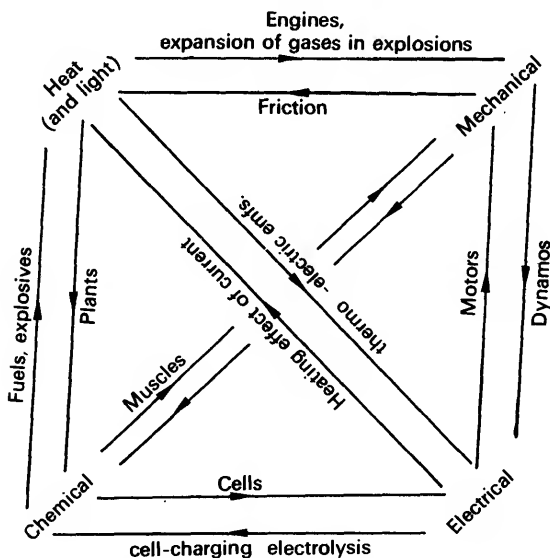


FIG. 8.7. Forms of energy, and their interconversions.

¹ DAVSON, *General Physiology*, Chap. XVII (Churchill).

² ABBOTT, AUBERT and HILL, *Jour. Physiology*, Vol. III.

from the sun—we would have no lakes if the sun did not evaporate the sea and provide the rainfall which fills the lakes. The relationship between all the principal forms of energy are summarized in Fig. 8.7.

Joule's Historic Experiments

About 1847 Joule measured the mechanical equivalent of heat by an apparatus of the form shown in Fig. 8.8. C is a copper cylinder, about 30 cm in diameter, containing water. The water is churned by paddles P, and prevented from whirling round *en masse* by baffles B. The paddles are connected by a coupling

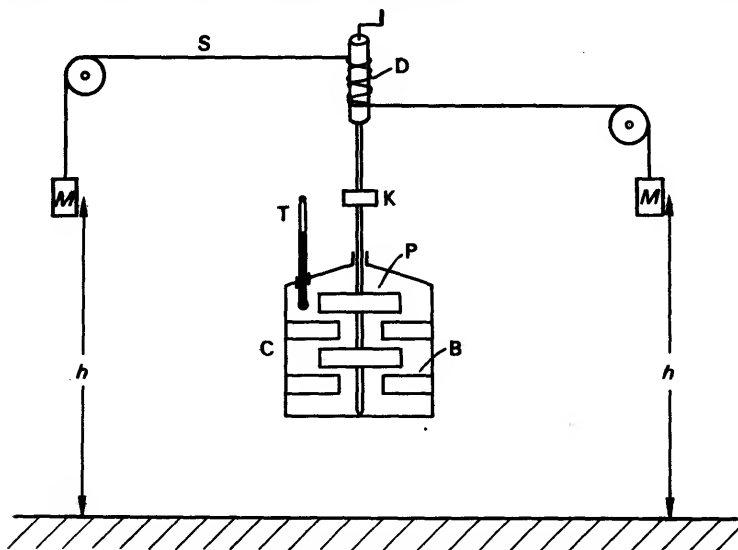


FIG. 8.8. Joule's apparatus for mechanical equivalent.

K to a drum D, which is rotated by strings S attached to lead weights M. A thermometer T shows the temperature of the water.

Joule would start an experiment by allowing the weights M to fall to the ground and turn the paddles. He would then break the coupling K, and re-wind the weights without disturbing the paddles. In this way he would make the weights fall twenty times or more, in a single experiment, and so increase the work done on the water and consequently the temperature rise.

Suppose n is the number of falls; then the work W' done on the water $= 2nMgh$, where M is the mass of one weight, g is the acceleration of gravity, and h is the height of the fall. The heat Q gained by the cylinder and the water is $(mc_w + C)\theta$, where m is the mass of water, c_w the heat capacity of the cylinder and paddles, and θ is their rise in temperature. The rise θ includes the correction for heat losses. The mechanical equivalent is given by W'/Q

$$= \frac{2nMgh}{(mc_w + C)\theta} \quad (1)$$

An experiment of this kind takes a long time—about half an hour—because a great deal of work, by everyday standards, must be done to produce a measurable amount of heat. The cooling correction is therefore relatively great.

Many people refused to accept Joule's work at first, because of the very small temperature differences on which it rested. Nevertheless, Joule's final result differs only by about one part in 400 from the value given by the best modern

experiments. In calculating his final result, Joule made corrections for the kinetic energy of the weights as they struck the floor, the work done against friction in the pulleys and the bearings of the paddle wheel, and the energy stored by the stretching of the strings; he even estimated the energy in the hum which the strings emitted, but found it was negligible.

Joule's Large-scale Experiments

In his last experiments, about 1878, Joule rotated the paddles with an engine,

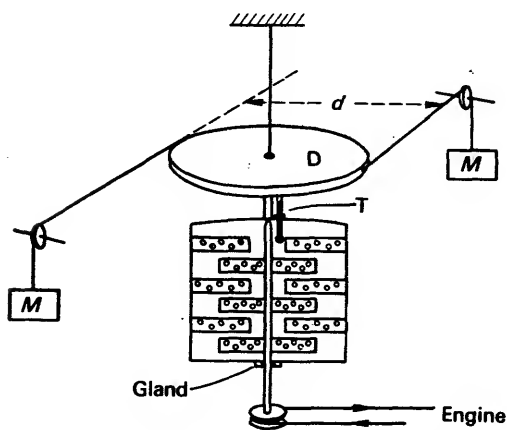


FIG. 8.9. Joule's final apparatus, also Rowland's.

is therefore

$$T = Mgd,$$

where d is the diameter of the wheel. Now the work done by a couple is equal to the product of its moment and the angle θ in radians through which it turns. Hence if the paddles make n revolutions, the work done on the water, since $\theta = 2\pi n$, is

$$\begin{aligned} W' &= 2\pi nT \\ &= 2\pi nMgd. \end{aligned}$$

The number of revolutions was measured on a revolution counter attached to the paddle spindle. If θ is the rise in temperature measured by the thermometer T, corrected for cooling, the heat developed is

$$Q = (mc_w + C)\theta,$$

in our previous notation. Hence the mechanical equivalent

$$= \frac{W'}{Q} = \frac{2\pi nMgd}{(mc_w + C)\theta}.$$

The term 'mechanical equivalent of heat', used in the past, has no meaning nowadays because heat is measured in joules in SI units (p. 194). Experiments in which mechanical energy is converted to heat energy are now regarded as experiments which measure the specific heat capacity of the heated substance in 'joule per kilogramme (or gramme) per deg K' ($\text{J kg}^{-1} \text{K}^{-1}$ or $\text{J g}^{-1} \text{K}^{-1}$). The specific heat capacity of water may be found by the energy conversion method described on p. 206.

thereby eliminating many of the corrections just mentioned. He suspended the cylinder on a wire, and kept it in equilibrium by an opposing couple, applied by means of the wheel D (Fig. 8.9). This method was repeated in 1880 by Rowland, who had holes drilled in the paddles and baffles, to make their churning action more thorough. The moment, T , of the couple applied by the engine is equal and opposite to that of the couple applied by the masses M . Its value

chapter nine

Calorimetry

Calorimetry is the measurement of heat; here we shall be concerned with the measurement of specific heat capacities and specific latent heats.

Heat (Thermal) Capacity, Specific Heat Capacity

The *heat capacity* of a body, such as a lump of metal, is the quantity of heat required to raise its temperature by 1 degree. It is expressed in *joules per deg K* (J K^{-1}).

The *specific heat capacity* of a substance is the heat required to warm unit mass of it through 1 degree; it is the heat capacity per unit mass of the substance. Specific heat capacities are expressed in *joule per kilogramme per deg K* ($\text{J kg}^{-1} \text{K}^{-1}$) or in *joule per gramme per deg K* ($\text{J g}^{-1} \text{K}^{-1}$). The specific heat of water, c_w , is about $4.2 \text{ J g}^{-1} \text{K}^{-1}$, or $4200 \text{ J kg}^{-1} \text{K}^{-1}$, or $4.2 \text{ kJ kg}^{-1} \text{K}^{-1}$, where $1 \text{ kJ} = 1 \text{ kilojoule} = 1000 \text{ J}$. Formerly, specific heat capacities were expressed in calories per gram per deg C—the values in joule are about 4.2 times as great.

From the definition of specific heat capacity, it follows that

$$\text{heat capacity, } C = \text{mass} \times \text{specific heat capacity.}$$

The specific heat capacity of copper, for example, is about $0.4 \text{ J g}^{-1} \text{K}^{-1}$ or $400 \text{ J kg}^{-1} \text{K}^{-1}$. Hence the heat capacity of 5 kg of copper = $5 \times 400 = 2000 \text{ J K}^{-1} = 2 \text{ kJ K}^{-1}$.

SPECIFIC HEAT CAPACITIES

Substance	Sp. Ht.: $\text{J kg}^{-1} \text{K}^{-1}$	Substance	Sp. Ht.: $\text{J kg}^{-1} \text{K}^{-1}$
Aluminium . . .	0.91×10^3	Ice	2.1×10^3
Brass	0.38	Paraffin wax . . .	2.9
Copper	0.39	Quartz	0.7
Iron	0.47	Rubber	1.7
Lead	0.13	Stone	0.9
Mercury	0.14	Wood.	1.7
Nickel	0.46	Alcohol	2.5
Platinum	0.13	Brine (25% by wt.)	3.4
Silver	0.24	Carbon tetra- chloride	0.84
Solder	0.18	Ether	2.4
Steel	0.45	Glycerine	2.5
Ebonite.	1.7	Paraffin oil	2.1
Glass	0.7	Turpentine	1.76

MEASUREMENT OF SPECIFIC HEAT CAPACITY

Method of Mixtures

A common way of measuring specific heat capacities is the method of mixtures, used by Wilcke (p. 193). Fig. 9.1 shows how we may

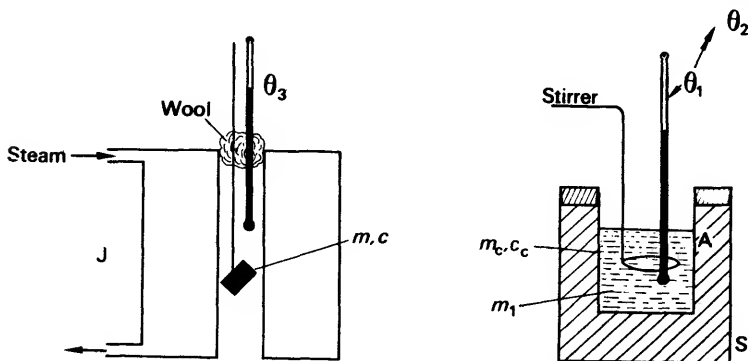


FIG. 9.1. Specific heat capacity by mixtures.

apply it to a solid, such as a metal. We weigh the specimen (m g) and hang it on a thread in a steam jacket, J , fitted with a thermometer. The jacket is plugged with cotton wool to prevent loss of heat by convection. While the solid is warming, we weigh a thin-walled copper vessel A called a calorimeter (m_c g), then run about 50 cm^3 of water into it, and by subtraction find the mass m_1 of this water. We put the calorimeter into a draught-shield S , and take the temperature, θ_1 , of the water in it. After we have given the specimen time to warm up—say half an hour—we read its temperature, θ_3 ; then we slide the calorimeter under the jacket, and drop the specimen into it. After stirring the mixture, we measure its final temperature, θ_2 . If no heat leaves the calorimeter by radiation, conduction, or convection, after the hot specimen has been dropped into the calorimeter, we have:

heat lost by solid in cooling from θ_3 to θ_2 = heat gained by water and calorimeter in warming from θ_1 to θ_2 .

(The heat gained by the thermometer and stirrer may be neglected if high accuracy is not required.)

Therefore, if c is the specific heat of the solid, c_w that of water and c_c that of the calorimeter:

$$\begin{aligned} mc(\theta_3 - \theta_2) &= m_1c_w(\theta_2 - \theta_1) + m_cc_c(\theta_2 - \theta_1) \\ &= (m_1c_w + m_cc_c)(\theta_2 - \theta_1), \end{aligned}$$

$$\text{whence } c = \frac{(m_1c_w + m_cc_c)(\theta_2 - \theta_1)}{m(\theta_3 - \theta_2)} \quad (1)$$

Liquids

The specific heat capacity of a liquid can be found by putting some in a calorimeter and dropping a hot solid, of known specific heat capacity c , into it. If m_p , c_l are the mass and specific heat capacity of the

liquid, then the product $m_i c_i$ replaces m in equation (1), from which c_i can be calculated.

Calculations

As an illustration of a specific heat capacity determination, suppose a metal of mass 200 g at 100°C is dropped into 80 g of water at 15°C contained in a calorimeter of mass 120 g and specific heat capacity 0.4 kJ kg⁻¹ K⁻¹. The final temperature reached is 35°C. Then:

$$\text{heat capacity of calorimeter} = 120 \times 0.4 = 48 \text{ J K}^{-1}$$

$$\text{heat capacity of water} = 80 \times 4.2 = 336 \text{ J K}^{-1}$$

$$\therefore \text{heat gained by water + cal.} = (336 + 48) \times (35 - 15) \text{ J}$$

$$\text{and heat lost by hot metal} = 0.2 \times c \times (100 - 35) \text{ J}$$

$$\therefore 0.2 \times c \times 65 = 384 \times 20$$

$$\therefore c = \frac{384 \times 20}{0.2 \times 65} = 590 \text{ J kg}^{-1} \text{ K}^{-1} \text{ (approx.)}$$

Heat Losses

In a calorimetric experiment, some heat is always lost by leakage. Leakage of heat cannot be prevented, as leakage of electricity can, by insulation, because even the best insulator of heat still has appreciable conductivity (p. 333).

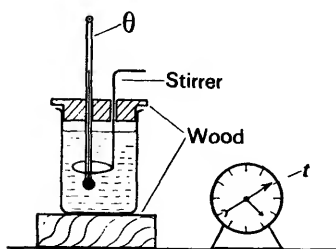
When convection is prevented, gases are the best thermal insulators. Hence calorimeters are often surrounded with a shield S, as in Fig. 9.1, and the heat loss due to conduction is made small by packing S with insulating material or by supporting the calorimeter on an insulating ring, or on threads. The loss by radiation is small at small excess temperatures over the surroundings. In some simple calorimetric experiments the final temperature of the mixture is reached quickly, so that the time for leakage is small. The total loss of heat is therefore negligible in laboratory experiments on the specific heats of metals, but not on the specific heat capacities of bad conductors, such as rubber, which give up their heat slowly. When great accuracy is required, the loss of heat by leakage is always taken into account.

Newton's Law of Cooling

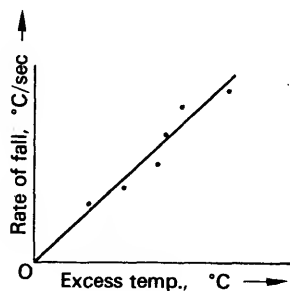
Newton was the first person to investigate the heat lost by a body in air. He found that *the rate of loss of heat is proportional to the excess temperature over the surroundings*. This result, called *Newton's law of cooling*, is approximately true in still air only for a temperature excess of about 20°C or 30°C; but it is true for all excess temperatures in conditions of forced convection of the air, i.e. in a draught. With natural convection Dulong and Petit found that the rate of loss of heat was proportional to $\theta^{5/4}$, where θ is the excess temperature, and this appears to be true for higher excess temperatures, such as from 50°C to 300°C. At low excess temperatures, however, less than 1°C,

G. T. P. Tarrant has pointed out that radiation, not convection, is the major contributing factor to the rate of cooling of an object.

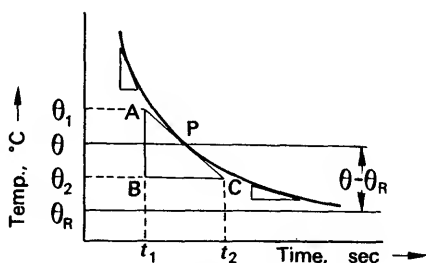
To demonstrate Newton's law of cooling, we plot a temperature (θ)-time (t) cooling curve for hot water in a calorimeter placed in a draught (Fig. 9.2 (a)). If θ_R is the room temperature, then the excess temperature of the water is $(\theta - \theta_R)$. At various temperatures, such as θ in Fig. 9.2 (b), we drew tangents such as APC to the curve. The slope of the tangent, in degrees per second, gives us the rate of fall of temperature, when the water is at the temperature θ :



(a) Apparatus



(c) Treatment of results



(b) Results

FIG. 9.2. Newton's law of cooling.

$$\text{rate of fall} = \frac{AB}{BC} = \frac{\theta_1 - \theta_2}{t_2 - t_1}$$

We then plot these rates against the excess temperature, $\theta - \theta_R$, as in Fig. 9.2 (c), and find a straight line passing through the origin. Since the heat lost per second by the water and calorimeter is proportional to the rate of fall of the temperature, Newton's law is thus verified.

Heat Loss and Temperature Fall

Besides the excess temperature, the rate of heat loss depends on the exposed area of the calorimeter, and on the nature of its surface: a dull surface loses heat a little faster than a shiny one, because it is a better radiator (p. 343). This can be shown by doing a cooling experiment twice, with equal masses of water, but once with the calorimeter polished, and once after it has been blackened in a candle-flame. In

general, for any body with a uniform surface at a uniform temperature θ , we may write, if Newton's law is true,

$$\text{heat lost/second} = \frac{dQ}{dt} = kS(\theta - \theta_R) \quad . \quad . \quad (2)$$

where S is the area of the body's surface, θ_R is the temperature of its surroundings, k is a constant depending on the nature of the surface, and O denotes the heat lost from the body.

When a body loses heat Q , its temperature θ falls; if m is its mass, and c its specific heat capacity, then its rate of fall of temperature, $d\theta/dt$, is given by

$$\frac{dQ}{dt} = -mc \frac{d\theta}{dt}.$$

Now the mass of a body is proportional to its volume. The rate of heat loss, however, is proportional to the surface area of the body. The rate of fall of temperature is therefore proportional to the ratio of surface to volume of the body. For bodies of similar shape, the ratio of surface to volume is inversely proportional to any linear dimension. If the bodies have surfaces of similar nature, therefore, the rate of fall of temperature is inversely proportional to the linear dimension: a small body cools faster than a large one. This is a fact of daily experience: a small coal which falls out of the fire can be picked up sooner than a large one; a tiny baby should be more thoroughly wrapped up than a grown man. In calorimetry by the method of mixtures, the fact that a small body cools faster than a large one means that, the larger the specimen, the less serious is the heat loss in transferring it from its heating place to the calorimeter. It also means that the larger the scale of the whole apparatus, the less serious are the errors due to loss of heat from the calorimeter.

Correction for Heat Losses in Calorimetry

Newton's law of cooling enables us to estimate the heat lost in an experiment on the method of mixtures.

In doing the experiment, we take the temperature of the mixture at half-minute intervals, and plot it against time, as in Fig. 9.3. The broken

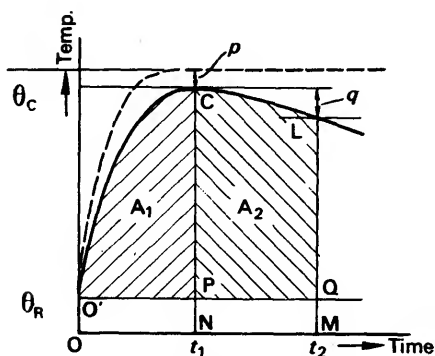


FIG. 9.3. Cooling correction.

line shows how we would expect the temperature to rise if no heat were lost; we have therefore to estimate the difference, p , between the plateau of this imaginary curve, and the crest of the experimental curve, C. p is known as the 'cooling correction'.

We start by drawing an ordinate CN through the crest, and another LM through any convenient point L further along the curve; OM should be not less than

twice ON—the greater it is, the more accurate the correction. We next draw an abscissa O'PQ through the room temperature, θ_R ; and by counting the squares of the graph paper, we measure the areas O'CP (A_1), PCLQ (A_2). Then, if q is the fall in temperature from C to L:

$$\frac{p}{q} = \frac{\text{O'CP}}{\text{PCLQ}} = \frac{A_1}{A_2}, \quad (3)$$

or

$$p = q \frac{A_1}{A_2}.$$

Before establishing this equation let us see how to use it. Suppose m_1, c , are the mass and specific heat capacity of the specimen; m, c_w are the mass and specific heat capacity of water; and C the heat capacity of the calorimeter. Then the heat which these lose to their surroundings is the heat which would have raised their temperature by p . Thus

$$\text{heat lost} = (m_1c + mc_w + C)p.$$

Let θ_1 denote the initial temperature of the specimen, θ_c the highest temperature of the mixture; and θ_2 the original temperature of the water and calorimeter. Then we have:

$$\text{heat given out} = \text{heat taken in} + \text{heat lost}.$$

$$\therefore m_1c(\theta_1 - \theta_c) = (mc_w + C)(\theta_c - \theta_2) + (m_1c + mc_w + C)p,$$

$$\text{from which } m_1c(\theta_1 - \overline{\theta_c + p}) = (mc_w + C)(\overline{\theta_c + p} - \theta_2).$$

To correct for the heat losses we must therefore add the correction p to the crest temperature θ on each side of the heat balance equation. In equation (1), p. 200, p must be added to θ_2 in both numerator and denominator.

Theory of the Correction. To establish equation (3), we write down the expression for the heat lost per second from the calorimeter, assuming Newton's law of cooling:

$$\frac{dQ}{dt} = kS(\theta - \theta_R) \quad (4)$$

where k is a constant, and S the exposed area of the calorimeter. Between times $t = 0$ and $t = t_1$, the total heat lost is

$$\begin{aligned} Q_1 &= \int_0^{t_1} kS(\theta - \theta_R) dt \\ &= kS \int_0^{t_1} (\theta - \theta_R) dt \\ &= kS \times \text{area O'CP} = kSA_1. \end{aligned}$$

This is the heat; which, if it had not been lost, would have warmed the calorimeter and contents by p degrees. Therefore

$$(m_1c + mc_w + C)p = kSA_1 \quad (5)$$

Similarly the heat lost between t_1 and t_2 is given by $Q_2 = kSA_2$, and since this loss caused a fall in temperature of q , we have, by the argument above

$$(m_1c + mc_w + C)q = kSA_2 \quad (6)$$

On dividing equation (5) by equation (6), we find

$$\frac{p}{q} = \frac{A_1}{A_2}, \text{ or } p = q \frac{A_1}{A_2}.$$

Specific Heat Capacity of Liquid by Cooling

Specific heat capacities of liquids which react with water are often measured by the so-called *method of cooling*. The cooling curve of a calorimeter is plotted, first when it contained a known volume of hot water, and then when it contains an *equal* volume of hot liquid (Fig. 9.4). The volumes are made equal so as to make the temperature distribution,

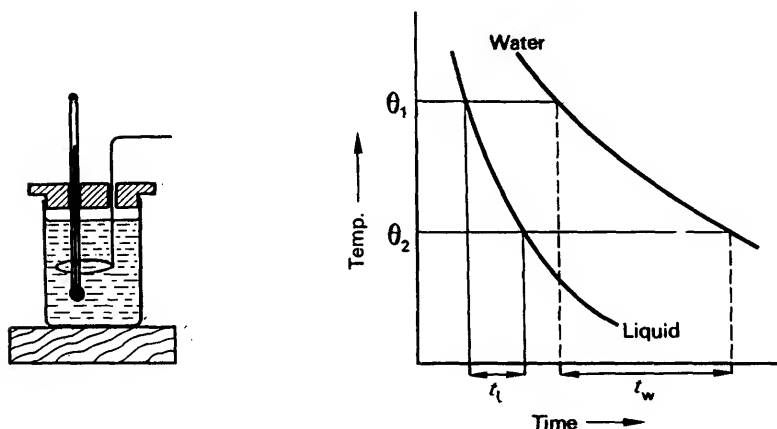


FIG. 9.4. Specific heat capacity by cooling.

over the surface of the calorimeter, the same in each experiment. From the curves, the respective times t_l and t_w are found which the calorimeter and contents take to cool from θ_1 to θ_2 . Whatever the contents of the calorimeter, it gives off heat at a rate which depends only on its excess temperature, since the area and nature of its surface are constant. Therefore, at each temperature between θ_1 and θ_2 , the calorimeter gives off heat at the same rate whatever its contents. Thus the average rate at which it loses heat, over the whole range, is the same with water and with liquid. Consequently

$$\frac{(m_1c + C)(\theta_1 - \theta_2)}{t_l} = \frac{(mc_w + C)(\theta_1 - \theta_2)}{t_w}$$

where m_1 , c , are the mass and specific heat capacity of the liquid, m , c_w that of water, and C is the heat capacity of the calorimeter. Thus

$$\frac{m_1c + C}{t_l} = \frac{mc_w + C}{t_w}$$

from which c can be calculated.

Specific Heat Capacity by Electrical Method

The simplest way to measure the specific heat capacity of a liquid in the laboratory is by electrical heating, as illustrated in Fig. 9.5. In this case, the energy supplied $= IVt$ joules, where I is the current in amperes the coil R of resistance wire, V is the potential difference across it in volts and t is the time in seconds for which current flows. The coil may be immersed in a suitable oil of mass m in a calorimeter of heat capacity C .

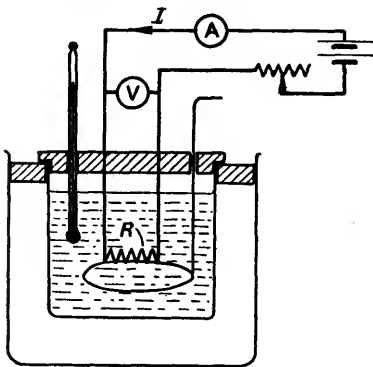


FIG. 9.5. Laboratory electrical method for specific heat capacity.

We pass a steady current I through the coil, and measure the potential difference V across it. Stirring continuously, we plot the temperature of the oil against the time. After a time t long enough to give several degrees rise, we switch off the current and plot the cooling curve. If θ is the corrected rise in temperature, we have

$$IVt = (mc + C)\theta,$$

whence we can calculate c in $\text{J kg}^{-1} \text{K}^{-1}$.

Specific Heat Capacity of Water by Continuous Flow Method

In 1899, Callendar and Barnes devised a method for specific heat capacity in which only steady temperatures are measured. They used platinum resistance thermometers, which are more accurate than mercury ones but take more time to read. In the measurement of steady temperatures, however, this is no drawback. As we shall see shortly *the heat capacity of the apparatus is not required*, which is a great advantage of the method.

Fig. 9.6 shows Callendar and Barnes' apparatus. Water from the constant-head tank K flows through the glass tube U , and can be collected as it flows out. It is heated by the spiral resistance wire R , which carries a steady electric current I . Its temperature, as it enters

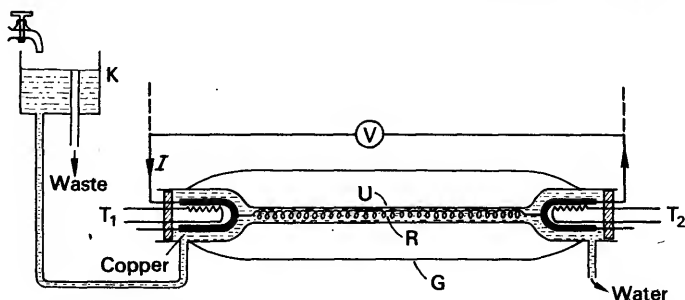


FIG. 9.6. Callendar and Barnes' apparatus (contracted several times in length relative to diameters).

and leaves, is measured by the thermometers T_1 and T_2 . (In a simplified laboratory experiment, these may be mercury thermometers.) Surrounding the apparatus is a glass jacket G , which is evacuated, so that heat cannot escape from the water by conduction or convection.

When the apparatus is running, it settles down eventually to a steady state, in which the heat supplied by the current is all carried away by the water. *None is then taken in warming the apparatus, because every part of it is at a constant temperature.* The mass of water m , which flows out of the tube in t seconds, is then measured. If the water enters at a temperature θ_1 and leaves at θ_2 , then if c_w is its mean specific heat capacity,

$$\text{heat gained by water} = Q = mc_w(\theta_2 - \theta_1) \text{ joules.}$$

The energy which liberates this heat is electrical. To find it, the current I , and the potential difference across the wire V , are measured with a potentiometer. If I and V are in amperes and volts respectively, then, in t seconds:

energy supplied to wire = IVt joules.

$$\therefore mc_w(\theta_2 - \theta_1) = IVt$$

$$\therefore c_w = \frac{IVt}{m(\theta_2 - \theta_1)}$$

To get the highest accuracy from this experiment, the small heat losses due to radiation, and conduction along the glass, must be allowed for. These are determined by the temperatures θ_1 and θ_2 . For a given pair of values of θ_1 and θ_2 , and constant-temperature surroundings (not shown), let the heat lost per second be h . Then, in t seconds,

$$\text{heat supplied by heating coil} = mc_w(\theta_2 - \theta_1) + ht,$$

$$\therefore IVt = mc_w(\theta_2 - \theta_1) + ht \quad (1)$$

To allow for the loss h , the rate of flow of water is changed, to about half or twice its previous value. The current and voltage are then adjusted to bring θ_2 back to its original value. The inflow temperature, θ_1 , is fixed by the temperature of the water in the tank. If I' , V' , are the new values of I , V , and m' is the new mass of water flowing in the same t seconds, then:

$$I'V't = m'c_w(\theta_2 - \theta_1) + ht \quad (2)$$

On subtracting equation (2) from equation (1), we find

$$(IV - I'V')t = (m - m')c_w(\theta_2 - \theta_1),$$

$$\therefore c_w = \frac{(IV - I'V')t}{(m - m')(\theta_2 - \theta_1)} \quad (3)$$

When the temperature rise, $\theta_2 - \theta_1$, is made small, for example, $\theta_1 = 20.0^\circ\text{C}$, $\theta_2 = 22.0^\circ\text{C}$, then c_w may be considered as the specific heat at 21.0°C , the mean temperature. If the inlet water temperature is now raised to say $\theta_1 = 40.0^\circ\text{C}$ and θ_2 is then 42.0°C , c_w is now the

specific heat at 41.0°C . In this way it was found that *the specific heat capacity of water varied with temperature*. The continuous flow method can be used to find the variation in specific heat capacity of any liquid in the same way.

The ' 15°C -calorie' was defined as the heat required to raise the temperature of 1 gramme of water from 14.5°C to 15.5°C . The table shows the relative variation of the specific heat capacity of water, taking the 15°C -calorie as 1.0000 in magnitude.

SPECIFIC HEAT CAPACITY OF WATER

Temperature ($^{\circ}\text{C}$)	5	15	25	40	70	100
c_w	1.0047	1.0000	0.9980	0.9973	1.0000	1.0057

Nernst's Method

Modern methods of measuring specific heat capacities use electrical heating. Nernst's method for the specific heat capacity of a metal is shown in Fig. 9.7. The metal S has a heating coil R of insulated platinum

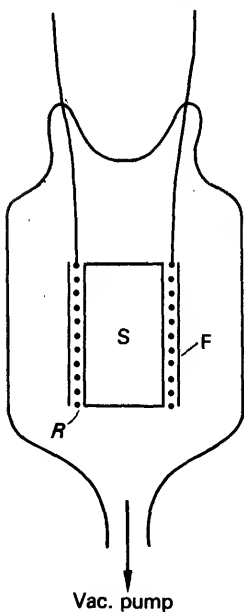


FIG. 9.7. Nernst's calorimeter.

wire wound round the outside, and is covered with silver foil F to minimize heat loss by radiation. It is suspended by the leads to the coil in a glass vessel, which is then evacuated, to prevent losses by convection and conduction. The resistance of the coil is measured, and from it the temperature is calculated. A steady current I , at a known potential difference V , is then passed through the coil for t seconds. After the current has been switched off, the resistance of the coil is again measured, to find the rise in temperature of the specimen. Resistance measurements are made at intervals, and enable the cooling curve to be plotted. If m is the mass of the specimen, C the heat capacity of the coil and foil, and θ the corrected rise in temperature, then the specific heat capacity c of the specimen is given by

$$IVt = (mc + C)\theta.$$

SPECIFIC LATENT HEAT

Fusion

The *specific latent heat of fusion* of a solid is the heat required to convert unit mass of it, at its melting-point, into liquid at the same temperature. It is expressed in joules per kilogramme (J kg^{-1}). High values can be more conveniently expressed in kJ kg^{-1} .

MELTING-POINTS AND SPECIFIC LATENT HEATS OF FUSION

Substance	M.P. °C	S.L.H. J kg ⁻¹	Substance	M.P. °C	S.L.H. J kg ⁻¹
Aluminium .	658	393 × 10 ³	Acetic acid . . .	17.5	184 × 10 ³
Antimony .	630	163	Beeswax	62	176
Bismuth . .	271	58	Brass	900	—
Copper . . .	1083	180	Naphthalene . .	80	146
Gold	1063	67	Paradichlorobenzene	53	—
Iron	1530	205	(non-flam.)		
Lead	327	25	Ice	0	333
Mercury . .	—38.9	12.5	Paraffin wax . .	50–60	—
Nickel . . .	1452	272	Solder (soft) . .	c. 180	50–85
Platinum . .	1773	113	Solder (hard) . .	c. 900	—
Sulphur . .	113–119	38	Steel	1400	—
Silver . . .	960.8	109	Glass	300–400	—
Sodium . . .	97.5	113	Quartz (fused) .	1,700	—
Tin	232	58	Hypo	48.2	c. 170
Tungsten . .	3380	—			
Zinc	419	109			

Ice is one of the substances whose specific latent heat of fusion we are likely to have to measure. To do so, place warm water, at a temperature θ_1 , a few degrees above room temperature, inside a calorimeter. Then add small lumps of ice, dried by blotting paper, until the temperature reaches a value θ_2 as much below room temperature as θ_1 was above. In this case a 'cooling correction' is not necessary. Weigh the mixture, to find the mass m of ice which has been added. Then the specific latent heat l is given by:

$$\left. \begin{array}{l} \text{heat given by calori-} \\ \text{meter and water in} \\ \text{cooling} \end{array} \right\} = \left\{ \begin{array}{l} \text{heat used in} \\ \text{melting ice} \end{array} \right\} + \left\{ \begin{array}{l} \text{heat used in warming} \\ \text{melted ice from } 0^\circ\text{C to} \\ \theta_2 \end{array} \right\}$$

$$\therefore (m_1 c_w + C)(\theta_1 - \theta_2) = ml + mc_w(\theta_2 - 0),$$

where m_1 = mass of water and c_w = specific heat capacity, C = thermal capacity of calorimeter, and θ_1 = initial temperature.

Hence
$$l = \frac{(m_1 c_w + C)(\theta_1 - \theta_2)}{m} - c_w \theta_2.$$

A modern electrical method, similar to Nernst's for specific heat capacities, gives

$$l = 334 \text{ kJ kg}^{-1} \quad \text{or} \quad 334 \text{ J g}^{-1}.$$

Bunsen's Ice Calorimeter

Bunsen's ice calorimeter is a device for measuring a quantity of heat by using it to melt ice.

When ice turns to water, it shrinks; the volume of 1 g of ice at 0°C is 1.0908 cm^3 , whereas that of 1 g of water at 0°C is 1.0001 cm^3 (p. 296). Thus the melting of 1 g of ice causes a contraction of 0.0907 cm^3 .

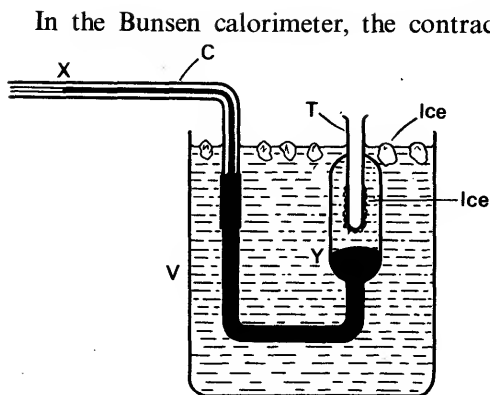


FIG. 9.8. Bunsen's ice calorimeter.

In the Bunsen calorimeter, the contraction due to the melting is measured, and from it the mass of ice melted is calculated. The apparatus is shown in Fig. 9.8. It consists of a test-tube T fused into a wider tube Y. The wider tube leads to a capillary C, and is filled with mercury from X to Y. The space above Y is filled with water from which all dissolved air has been boiled.

Except for the capillary, the whole apparatus is placed in ice-water in the vessel V, and, after some time, it all settles down to 0°C . A little ether is then poured into T, and air is blown through it via a thin tube; the ether evaporates and cools the tube T, so that ice forms on the outside of it. A pad of cotton wool is then dropped to the bottom of T and the apparatus is left for some more time, to allow the newly formed ice to settle down to 0°C .

When the apparatus is ready for use, the end of the mercury thread in C is observed by a travelling microscope. If the specific heat capacity of a solid is to be measured, the specimen is weighed (m g) and left to come to room temperature θ . The solid is then gently dropped into the tube T. As it cools, it melts ice, and causes the mercury thread to run back along the capillary. When the thread has ceased to move, its end is again observed. If it has moved through l cm, and the cross-section of the capillary is a cm^2 , then the contraction is al cm^3 . The mass of ice melted is therefore $al/0.0907$ g, and the heat absorbed is $334 al/0.0907$ joules. This heat is given out by m g of solid cooling from θ to 0°C ; the specific heat capacity c of the solid is therefore given by

$$mc\theta = \frac{334 al}{0.0907}.$$

In practice, the cross-section is not measured, and the instrument is calibrated by dropping into it a solid of known mass, m_1 , and specific heat capacity, c_1 . If the room temperature is constant, then

$$\frac{mc}{m_1 c_1} = \frac{l}{l_1},$$

where l_1 is the displacement of the mercury in the calibration experiment. Thus c can be found.

Advantages of the Ice Calorimeter

The advantages of the ice calorimeter are:

- (i) no correction for heat capacity of the container: the specimen tube starts at 0°C and finishes at 0°C —all the heat from the specimen is used to melt ice, at constant temperature;

- (ii) no heat losses from the apparatus—it is surrounded by a bath at the same temperature as itself, and therefore neither loses heat to the outside, nor gains any from it;
- (iii) no loss of heat from the specimen before it enters the calorimeter—the specimen starts at room temperature, and therefore gives up no heat until it enters the specimen tube (contrast the method of mixtures, in which the specimen is heated to 100°C or so): this is a great advantage when the specimen is small;
- (iv) easy, and therefore accurate, thermometry—the only temperature to be measured is the room temperature, which is constant and can be determined at leisure.

An advantage sometimes asserted is that specimens can be added one after another, without having to re-set the apparatus. That is true, because each specimen comes to 0°C in turn, and then behaves simply like part of the apparatus, taking no heat from any following specimen. But it does not mean that the calorimeter has the advantage of speed—the time taken to set it up would be enough for half a dozen measurements by the method of mixtures. A disadvantage of this calorimeter is that it never settles down completely—the mercury is always slowly creeping along the capillary, and the creep during an experiment must be estimated and allowed for.

The calorimeter was devised in 1871; it is rarely used nowadays, because electrical methods of calorimetry are more convenient and accurate. However, it has been used for measuring the specific heat capacities of rare earths of which only small specimens were available.

Evaporation

The *specific latent heat of evaporation* of a liquid is the heat required to convert unit mass of it, at its boiling-point, into vapour at the same temperature. It is expressed in joule per kilogramme (J kg^{-1}), or, with high values, in kJ kg^{-1} .

BOILING-POINTS AND SPECIFIC LATENT HEATS OF EVAPORATION

Substance	B.P. °C	S.L.H. J kg^{-1}
Aluminium	1800	—
Acetone	56.7	—
Alcohol (ethyl)	78.3	867×10^3
Alcohol (methyl)	64.7	1120
Benzene	80.2	389
Carbon disulphide	46.2	351
Carbon tetrachloride (non-flam.)	76.7	193
Ether	34.6	370
Glycerine	290	—
Turpentine	161	—
Mercury	357	272
Platinum	3910	—
Sodium	877	—
Sulphur	444.6	—

To find the specific latent heat of evaporation of water, we pass steam into a calorimeter with water (Fig. 9.9). On its way the steam passes through a vessel, T in the figure, which traps any water carried over by the steam and is called a steam-trap. The mass m of condensed steam is found by weighing. If θ_1 and θ_2 are the initial and final temperatures of the water, the specific latent heat l is given by:

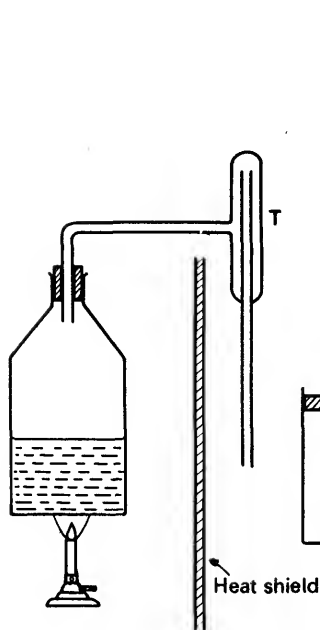


FIG. 9.9. Latent heat of evaporation of water.

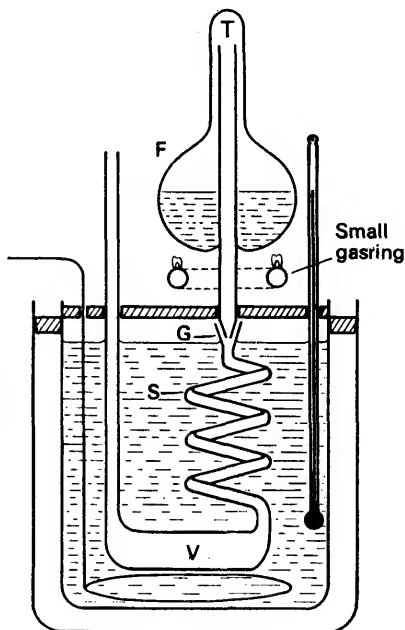


FIG. 9.10. Berthelot's apparatus for latent heat of evaporation.

$$\left. \begin{array}{l} \text{heat given by steam} \\ \text{condensing} \end{array} \right\} + \left\{ \begin{array}{l} \text{heat given by con-} \\ \text{densed water cooling} \\ \text{from } 100^{\circ}\text{C to } \theta_2 \end{array} \right\} = \left\{ \begin{array}{l} \text{heat taken by} \\ \text{calorimeter and} \\ \text{water} \end{array} \right.$$

$$ml + mc_w(100 - \theta_2) = (m_1c_w + C)(\theta_2 - \theta_1)$$

where m_1c_w and C have their usual meanings.

Hence
$$l = \frac{(m_1c_w + C)(\theta_2 - \theta_1)}{m} - c_w(100 - \theta_2).$$

The accepted value of the specific latent heat of evaporation of water is about $l = 2260 \text{ kJ kg}^{-1}$ or 2260 J g^{-1} .

Berthelot's Apparatus. An apparatus suitable for use with liquids other than water was devised by Berthelot in 1877 (see Fig. 9.10). The liquid is boiled in the flask F, and its vapour passes out through the tube T. This fits with a ground joint G into the glass spiral S, which is surrounded by water in a calorimeter. The vapour condenses in the

spiral, and collects in the vessel V, where it can afterwards be weighed.
Let

θ_b = boiling-point of liquid.

c = specific heat capacity of liquid.

m = mass of liquid condensed.

θ_1 = initial temperature of water.

θ_2 = final temperature of water, *corrected for cooling*.

m_1 = mass of water of specific heat c_w .

C = thermal capacity of calorimeter + glassware below joint.

Then $ml + mc(\theta_b - \theta_2) = (m_1 c_w + C)(\theta_2 - \theta_1)$,

whence $l = \frac{(m_1 c_w + C)(\theta_2 - \theta_1)}{m} - c(\theta_b - \theta_2)$.

Electrical Method for Specific Latent Heat

A modern electrical method for the specific latent heat of evaporation of water is illustrated in Fig. 9.11 below. The liquid is heated in a vacuum-jacketed vessel U by the heating coil R. Its vapour passes down

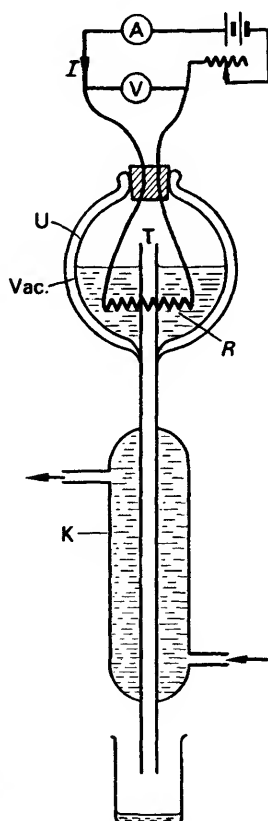


FIG. 9.11. Electrical method for latent heat of evaporation.

the tube T, and is condensed by cold water flowing through the jacket K. When the apparatus has reached its steady state, the liquid is at its boiling-point, and the heat supplied by the coil is used in evaporating the liquid, and in offsetting the losses. The liquid emerging from the condenser is then collected for a measured time, and weighed.

If I and V are the current through the coil, and the potential difference across it, the electrical energy supplied in t seconds is IVt . And if h is the heat lost from the vessel per second, and m the mass of liquid collected in t seconds, then

$$IVt = ml + ht \quad (1)$$

The heat losses h are determined by the temperature of the vessel, which is fixed at the boiling-point of the liquid. Therefore they may be eliminated by a second experiment with a different rate of evaporation (cf. Callendar and Barnes, p. 206). If $I'V'$ are the new current and potential difference, and if m' grammes are evaporated in t seconds, then

$$I'V't = m'l + ht.$$

Hence by subtraction from equation (1)

$$l = \frac{(IV - I'V')t}{(m - m')}.$$

EXAMPLES

1. An electric kettle has a 750 W–240 V heater and is used on a 200 V mains. If the heat capacity of the kettle is 400 J K^{-1} and the initial water temperature is 20°C , how long will it take to boil 500 g of water, assuming the resistance of the heater is unaltered on changing to the new mains.

Firstly, find the new power absorbed on the 200 V mains. Since the resistance R is constant and $P = V^2/R$, it follows that $P \propto V^2$.

$$\therefore \text{new power} = \left(\frac{200}{240}\right)^2 \times 750 \text{ W} = 520 \text{ W (approx.)}$$

$$\therefore \text{heat supplied to water} = 520 \text{ J per second} \quad (1)$$

Secondly, assuming 100°C is the boiling point and $4.2 \text{ kJ kg}^{-1} \text{ K}^{-1}$ ($4.2 \text{ J g}^{-1} \text{ K}^{-1}$) is the specific heat capacity of water,

$$\begin{aligned} \text{heat gained by water and kettle} &= 500 \times 4.2 \times (100 - 20) + 400 \times (100 - 20) \\ &= (500 \times 4.2 + 400)(100 - 20) \\ &= 610 \times 80 \text{ J.} \end{aligned}$$

$$\text{From (1), } \therefore \text{time, } t = \frac{2500 \times 80}{520} = 385 \text{ seconds (approx.)} = 6.4 \text{ min.}$$

2. Water flows at the rate of 150.0 g min^{-1} through a tube and is heated by a heater dissipating 25.2 W. The inflow and outflow water temperatures are 15.2°C and 17.4°C respectively. When the rate of flow is increased to 231.8 g min^{-1} and the rate of heating to 37.8 W, the inflow and outflow temperatures are unaltered. Find (i) the specific heat capacity of water, (ii) the rate of loss of heat from the tube.

Suppose c_w is the specific heat of water in $\text{J g}^{-1} \text{K}^{-1}$ and h is the heat lost in J s^{-1} . Then, since $1 \text{ W} = 1 \text{ J per second}$,

$$25.2 = \frac{150.0}{60} c_w (17.4 - 15.2) + h \quad (1)$$

and
$$37.8 = \frac{231.8}{60} c_w (17.4 - 15.2) + h \quad (2)$$

Subtracting (1) from (2),

$$\begin{aligned} \therefore 37.8 - 25.2 &= \frac{231.8 - 150.0}{60} c_w (17.4 - 15.2) \\ \therefore c_w &= \frac{12.6 \times 60}{81.8} = 4.2 \text{ J g}^{-1} \text{K}^{-1} = 4.2 \text{ kJ kg}^{-1} \text{K}^{-1}. \end{aligned}$$

Substituting for c_w in (1),

$$\therefore h = 25.2 - \frac{150}{60} \times 4.2 \times 2.2 = 2.1 \text{ J s}^{-1}.$$

3. Define latent heat. Describe the measurement of the specific latent heat of evaporation of water under school laboratory conditions.

A copper calorimeter of mass 70.5 g contains 100.0 g of water at 19.5°C . Naphthalene (M.P. 79.9°C) is melted in a test tube, cooled to 80.0°C , and then poured into the calorimeter. If the highest temperature reached by the water after stirring is 28.7°C and the final mass of the calorimeter and its contents is 188.3 g calculate the latent heat of fusion of naphthalene. (Specific heat capacity of copper 0.4 , of naphthalene $1.3 \text{ kJ kg}^{-1} \text{K}^{-1}$.) (L.)

First part. The specific latent heat of a substance is the heat required to change unit mass of the solid at the melting-point to liquid at the same temperature (fusion), or the heat required to change unit mass of the liquid at the boiling-point to vapour at the same temperature (vaporization).

The measurement of the specific latent heat of evaporation of water requires the following, among other points: (i) use of a steam trap, (ii) a rise in temperature of the water in the calorimeter of about 10°C , (iii) a 'correction' to 100°C as the steam temperature, if the barometric pressure is not 76 cm mercury, (iv) a cooling correction.

Second part. The mass of naphthalene = $188.3 - 170.5 = 17.8 \text{ g}$.

Heat lost by naphthalene = heat gained by water and calorimeter.

$$\begin{aligned} \therefore 17.8l + 17.8 \times 1.3 \times (79.9 - 28.7) \\ = 100 \times 4.2 \times (28.7 - 19.5) + 70.5 \times 0.4 \times (28.7 - 19.5). \end{aligned}$$

Solving,
$$\therefore l = 164 \text{ J g}^{-1} \text{ (approx.)} = 164 \text{ kJ kg}^{-1}.$$

4. In an X-ray tube, 10^{18} electrons per second arrive with a speed of $2 \times 10^6 \text{ m s}^{-1}$ at a metal target of mass 200 g and specific heat capacity $0.5 \text{ J g}^{-1} \text{K}^{-1}$. If the mass of an electron is $9.1 \times 10^{-31} \text{ kg}$, and assuming 98% of the incident energy is converted into heat, find how long the target will take to rise in temperature by 50°C assuming no heat losses.

The kinetic energy of a moving object is $\frac{1}{2}mv^2$ joules, where m is the mass in kg and v is the speed in m s^{-1} . Assuming the initial speed is zero,

$$\begin{aligned} \therefore \text{energy per second of incident electrons} &= \frac{1}{2} \times (10^{18} \times 9.1 \times 10^{-31}) \times (2 \times 10^6)^2 \text{ J} \\ &= 1.8 \text{ J (approx.)} \end{aligned}$$

$$\text{Heat gained by target} = 200 \times 0.5 \times 50 \text{ J} = 50000 \text{ J}$$

$$\therefore \text{time} = \frac{50000}{1.8} = 2780 \text{ seconds (approx.)} = 46.3 \text{ min.}$$

EXERCISES 9

1. Describe, with the aid of a labelled diagram, how you would find the specific heat capacity of a liquid by the method of continuous flow.

Discuss the advantages and disadvantages of the method compared with the method of mixtures.

The temperature of 50 g of a liquid contained in a calorimeter is raised from 15.0°C (room temperature) to 45.0°C in 530 seconds by an electric heater dissipating 10.0 watts. When 100 g of liquid is used and the same change in temperature occurs in the same time, the power of the heater is 16.1 watts. Calculate the specific heat capacity of the liquid. (*N.*)

2. Distinguish between *specific* heat capacity and *latent* heat capacity. With what physical changes is each associated? Describe the processes involved in terms of simple molecular theory.

A thin-walled tube containing 5 cm^3 of ether is surrounded by a jacket of water calibrated so that changes in the volume of the water can be read off. The whole apparatus is cooled down to 0°C and all the ether is then evaporated by blowing a rapid stream of air pre-cooled to 0°C through it. The change of volume as ice forms in the water is 0.35 cm^3 . Calculate the specific latent heat of evaporation of the ether.

(Use the following values: specific latent heat of fusion of ice = 334 J g^{-1} . Densities at 0°C : water, 1.000 g cm^{-3} ; ice, 0.917 g cm^{-3} ; ether, 0.736 g cm^{-3} .) (*O. & C.*)

3. Give an account of an electrical method of finding the specific latent heat of vaporisation of a liquid boiling at about 60°C . Point out any causes of inaccuracy and explain how to reduce their effect.

Ice at 0°C is added to 200 g of water initially at 70°C in a vacuum flask. When 50 g of ice has been added and has all melted the temperature of the flask and contents is 40°C . When a further 80 g of ice has been added and has all melted the temperature of the whole becomes 10°C . Calculate the specific latent heat of fusion of ice, neglecting any heat lost to the surroundings.

In the above experiment the flask is well shaken before taking each temperature reading. Why is this necessary? (*C.*)

4. The specific heat capacity of gallium metal is $0.33\text{ kJ kg}^{-1}\text{ K}^{-1}$. Explain carefully how this result may be determined experimentally. Indicate the sources of error in your method and estimate the accuracy which could be achieved.

[Melting point of gallium = 30°C .]

A ball of gallium is released from a stationary balloon, falls freely under gravity and on striking the ground it just melts. Calculate the height of the balloon assuming that the temperature of the gallium just before impact is 1°C and that all the energy gained during its free fall is used to heat the gallium on impact. Why are the conditions specified in this problem unrealistic?

[Specific latent heat of fusion of gallium = 79 kJ kg^{-1} .] (*O. & C.*)

5. Explain what is meant by the specific latent heat of vaporization of a liquid, and describe an experiment for an accurate determination of this quantity for carbon tetrachloride, which boils at 77°C .

A thermally insulated vessel connected to a vacuum pump contains 10.0 g of water at a temperature of 0°C . As air and water vapour are exhausted from the vessel, it is observed that the water remaining in the vessel freezes. Explain why this happens, and find the mass of water which is converted into ice.

[Specific latent heat of vaporization of water at 0°C = 2520 kJ kg^{-1} ; specific latent heat of fusion of ice at 0°C = 336 kJ kg^{-1} .] (*C.*)

6. Discuss the nature of the heat energy (a) of a solid, (b) of a gas, (c) of the sun during the transmission of this energy to the earth.

265430 joules of heat are produced when a vehicle of total mass 1270 kg is brought to rest on a level road. Calculate the speed of the vehicle in km per hr just before the brakes are applied. (L.)

7. State *Newton's law of cooling* and describe how to obtain observations and how to use them in order to test the validity of the law.

A solid of mass 250 g in a vessel of thermal capacity 67.2 J K^{-1} is heated to a few degrees above its melting point and allowed to cool in steady conditions until solid again. Sketch the graph of its temperature plotted against time. If this graph shows that immediately before solidification starts the rate of cooling is $3.2 \text{ deg C min}^{-1}$, while immediately after solidification it is $4.7 \text{ deg C min}^{-1}$, calculate the specific heat capacity of the solid taken by the solidifying process. (The specific latent heat of fusion of the substance may be taken as 146.6 kJ kg^{-1} and its specific heat capacity in the liquid state as $1.22 \text{ kJ kg}^{-1} \text{ K}^{-1}$.) (L.)

8. What do you understand by the *specific heat capacity* of a substance? Describe how you would measure the specific heat capacity of a sample of rock, describing the precautions that you would take to obtain an accurate result.

A room is heated during the day by a 1 kW electric fire. The fire is to be replaced by an electric storage heater consisting of a cube of concrete which is heated overnight and is allowed to cool during the day, giving up its heat to the room. Estimate the length of an edge of the cube if the heat it gives out in cooling from 70°C to 30°C is the same as that given out by the electric fire in 8 hours.

[Density of concrete = 2700 kg m^{-3} ; specific heat capacity of concrete = $0.85 \text{ kJ kg}^{-1} \text{ K}^{-1}$.] (O. & C.)

9. An experiment was performed to determine the specific latent heat of vaporization of a volatile liquid at the prevailing boiling point by the method of electrical heating. The results are summarized in the following table:

Rate of supply of energy to boiling liquid (watt)	Mass of liquid vaporized in 200 seconds (g)
10	1.6
20	6.4
30	11.2
40	16.0

Use the data to plot a graph (using the graph paper available) and hence determine the latent heat of vaporization of the liquid and the rate of loss of heat from the calorimeter containing the boiling liquid.

Draw a labelled diagram of a suitable apparatus for use in the experiment and indicate how the above results would have been obtained.

Give two advantages of this method over the method of mixtures. (N.)

10. In 1948 the International Conference on Weights and Measures recommended that the calorie should no longer be regarded as the basic unit of heat, but that it should be replaced by the joule. Discuss the reasons for, and the advantages and possible disadvantages of, this recommendation.

Give a labelled diagram of a continuous flow calorimeter suitable for the determination of the specific heat capacity of a liquid. What measurements would you make in such a determination, and how would the result be obtained from them? State the precautions which you would adopt to ensure an accurate result. (C.)

11. Give an account of a method of determining the specific latent heat of evaporation of water, pointing out the ways in which the method you describe achieves, or fails to achieve, high accuracy.

A 600 watt electric heater is used to raise the temperature of a certain mass of water from room temperature to 80°C . Alternatively, by passing steam from a boiler into the same initial mass of water at the same initial temperature the same temperature rise is obtained in the same time. If 16 g of water were being evaporated every minute in the boiler, find the specific latent heat of steam, assuming that there were no heat losses. (O. & C.)

12. 300 g of a certain metal of density about 19 g cm^{-3} is available in the form of a coarse powder, together with a calorimeter of heat capacity about 33.6 J K^{-1} and volume about 160 cm^3 , and a 50°C thermometer reading to $\frac{1}{5}\text{ deg}$.

Using this and other necessary apparatus, how would you verify, by the method of mixtures, that the specific heat of the metal is $0.13\text{ kJ kg}^{-1}\text{ K}^{-1}$?

In the experiment you describe why is it (a) unnecessary to apply a correction for heat exchange with the surroundings, (b) necessary to decide on a suitable maximum temperature of the mixture? How would you ensure that such a temperature is realized? (N.)

13. Describe an electrical method for the determination of the specific latent heat of steam. State the probable sources of error in the experiment and suggest how they may be minimised.

In one method for storing solar energy, Glauber's salt can be allowed to warm up to 45°C in the sun's rays during the day and the stored heat is used during the night, the salt cooling down to 25°C . Glauber's salt melts at 32°C . Calculate the mass of salt needed to store 1 million joules.

(Specific heat capacity of Glauber's salt, solid = $0.11\text{ kJ kg}^{-1}\text{ K}^{-1}$; specific heat capacity of Glauber's salt, molten = $0.16\text{ kJ kg}^{-1}\text{ K}^{-1}$. Specific latent heat of fusion = 14 kJ kg^{-1} .) (O. & C.)

14. What is meant by the specific heat capacity of a substance? Give a brief account of two methods, one in each case, which may be used to find the specific heat capacity of each of the following: (a) a specimen of a metal in the form of a block a few cm in linear dimensions, and (b) a liquid which is available in large quantities. Indicate whether the methods you describe involve a cooling correction.

An electrical fuse consists of a piece of lead wire 1.5 mm in diameter and 5 cm long, which has a resistance of $6.5 \times 10^{-3}\text{ ohm}$. Owing to a fault a constant current of 800 A passes through the fuse. If the wire is initially at 10°C and melts at 330°C , find the time interval before it starts to melt, assuming that its specific heat capacity and its electrical resistance are constant and that there are no heat losses.

[Specific heat capacity of lead = $0.13\text{ kJ kg}^{-1}\text{ K}^{-1}$. Density of lead = 11000 kg m^{-3} .] (O. & C.)

15. Describe the determination of the latent heat of fusion of ice by the method of mixtures and, in particular, show how allowance is made for heat interchange with the surroundings.

A calorimeter of heat capacity 84 J K^{-1} contains 980 g of water supercooled to -4°C . Taking the latent heat of fusion of ice at 0°C as 336 kJ kg^{-1} , find the amount of ice formed when the water suddenly freezes. Calculate also the specific latent heat of fusion at -4°C if the specific heat capacity of ice is $2.1\text{ kJ kg}^{-1}\text{ K}^{-1}$. (N.)

16. State Newton's law of cooling, and describe an experiment by which you would verify it. A calorimeter containing first 40 and then 100 g of water is heated and suspended in the same constant-temperature enclosure. It is found

that the times taken to cool from 50° to 40°C in the two cases are 15 and 33 minutes respectively. Calculate the heat capacity of the calorimeter. (O. & C.)

17. Oil at 15.6°C enters a long glass tube containing an electrically heated platinum wire and leaves it at 17.4°C , the rate of flow being 25 cm^3 per min and the rate of supply of energy 1.34 watts. On changing the rate of flow to 15 cm^3 per min and the power to 0.76 watt the temperature again rises from 15.6° to 17.4°C . Calculate the mean specific heat capacity of the oil between these temperatures. Assume that the density of the oil is 870 kg m^{-3} . (N.)

18. In the absence of bearing friction a winding engine would raise a cage weighing 1000 kg at 10 m s^{-1} , but this is reduced by friction to 9 m s^{-1} . How much oil, initially at 20°C , is required per second to keep the temperature of the bearings down to 70°C ? (Specific heat capacity of oil = $2.1\text{ kJ kg}^{-1}\text{ K}^{-1}$; $g = 9.81\text{ m s}^{-2}$. (O. & C.)

19. A heating coil is embedded in a copper cylinder which also carries a thermocouple. The whole is thermally equivalent to 25 g of copper. The cylinder is suspended in liquid air until the thermocouple reading is constant. The cylinder is taken out and rapidly transferred into a beaker of water at 0°C . A coating of ice forms on the cylinder and when its temperature is again constant it is taken out of the water and suspended in a space maintained at 0°C . The heating coil is switched on at a steady energy dissipation of 24 watts. After 1 minute 5 seconds the whole of the ice has just melted. What is the temperature of the liquid air?

What assumptions were made in carrying out the calculations? (Mean specific heat capacity of copper is $0.336\text{ kJ kg}^{-1}\text{ K}^{-1}$.) (L.)

20. Describe a continuous flow method of measuring the specific heat capacity of a liquid. Explain the advantages of the method.

Use the following data to calculate the specific heat capacity of the liquid flowing through a continuous flow calorimeter: *Experiment I*. Current 2.0 amp, applied p.d. 3.0 volt, rate of flow of liquid 30 g min^{-1} , rise in temperature of liquid 2.7°C . *Experiment II*. Current 2.5 amp, applied p.d. 3.75 volts, rate of flow of liquid 48 g min^{-1} , rise in temperature of liquid 2.7°C . (L.)

chapter ten

Gases

IN this chapter we shall be concerned with the relationship between the temperature, pressure and volume of a gas. Unlike the case of a solid or liquid this can be expressed in very simple laws, called the Gas Laws, and reduced to a simple equation, called the Equation of State. We shall also deal in this chapter with the specific heat capacities of gases.

THE GAS LAWS AND THE EQUATION OF STATE

Pressure and Volume: Boyle's Law

In 1660 Robert Boyle—whose epitaph reads ‘Father of Chemistry, and Nephew of the Earl of Cork’—published the results of his experiments on the natural spring of air. In the vigorous language of the seventeenth century, he meant what we now tamely call the relationship between the pressure of air and its volume. Similar results were published in 1676 by Mariotte, who had not heard of Boyle's work. Boyle trapped air in the closed limb of a U-tube, with mercury (Fig. 10.1 (a)). He first adjusted the amount of mercury until its level was the same in each limb, so that the trapped air was at atmospheric pressure. He next

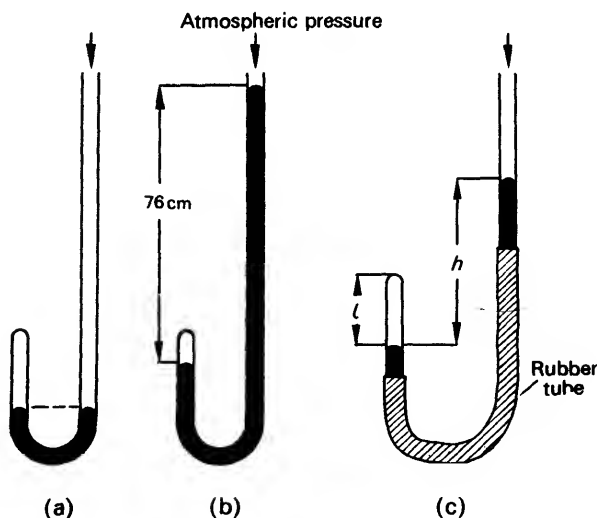


FIG. 10.1. Boyle's law apparatus.

poured in more mercury, until he had halved the volume of the trapped air (Fig. 10.1 (b)). Then 'not without delight and satisfaction' he found that the mercury in the open limb stood 740 mm above the mercury in the closed limb. Since he knew that the height of the barometer was about 740 mm of mercury, he realized that to halve the volume of his air he had had to double the pressure on it.

We can repeat Boyle's experiment with the apparatus shown in Fig. 10.1 (c); its form makes the pouring of mercury unnecessary. We set the open limb of the tube at various heights above and below the closed limb and measure the difference in level, h , of the mercury. When the mercury in the open limb is below that in the closed, we reckon h as negative. At each value of h we measure the corresponding length l of the air column in the closed limb. To find the pressure of the air we add the difference in level h to the height of the barometer, H ; their sum gives the pressure p of the air in the closed limb:

$$p = g\rho(H + h)$$

where g is the acceleration of gravity and ρ is the density of mercury.

If S is the area of cross-section of the closed limb, the volume of the trapped air is

$$V = lS.$$

To interpret our measurements we may either plot $H+h$, which is a measure of p against $1/l$ or tabulate the product $(H+h)l$. We find that the plot is a straight line, and therefore

$$(H+h) \propto \frac{1}{l} \quad . \quad . \quad . \quad . \quad (1)$$

Alternatively, we find

$$(H + h)l = \text{constant}, \quad (2)$$

which means the same as (1).

Since, g , ρ , and S are constants, the relationships (1) and (2) give

$$p \propto \frac{1}{V}$$

or

$pV = \text{constant.}$

A little later in this chapter we shall see that the pressure of a gas depends on its temperature as well as its volume. To express the results of the above experiments, therefore, we say that *the pressure of a given mass of gas, at constant temperature, is inversely proportional to its volume*. This is *Boyle's Law*.

Mixture of Gases: Dalton's Law

Fig. 10.2 shows an apparatus with which we can study the pressure of a mixture of gases. A is a bulb, of volume V_1 , containing air at atmospheric pressure, p_1 . C is another bulb, of volume V_2 , containing

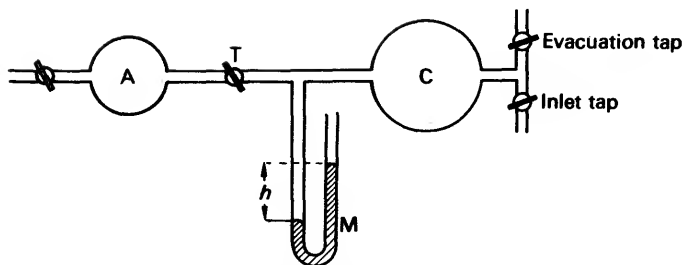


FIG. 10.2. Apparatus for demonstrating Dalton's law of partial pressures.

carbon dioxide at a pressure p_2 . The pressure p_2 is measured on the manometer M; in millimetres of mercury it is

$$p_2 = h + H,$$

where H is the height of the barometer. (In the same units, the air pressure, $p_1 = H$.)

When the bulbs are connected by opening the tap T, the gases mix, and reach the same pressure, p ; this pressure is given by the new height of the manometer. Its value is found to be given by

$$p = p_1 \frac{V_1}{V_1 + V_2} + p_2 \frac{V_2}{V_1 + V_2}.$$

Now the quantity $p_1 V_1 / (V_1 + V_2)$ is the pressure which the air originally in A would have, if it expanded to occupy A and C; for, if we denote this pressure by p' , then $p'(V_1 + V_2) = p_1 V_1$. Similarly $p_2 V_2 / (V_1 + V_2)$ is the pressure which the carbon dioxide originally in C would have, if it expanded to occupy A and C. Thus the total pressure of the mixture is the sum of the pressures which the individual gases exert, when they have expanded to fill the vessel containing the mixture.

The pressure of an individual gas in a mixture is called its *partial pressure*: it is the pressure which would be observed if that gas alone occupied the volume of the mixture, and had the same temperature as the mixture. The experiment described shows that *the pressure of a mixture of gases is the sum of the partial pressures of its constituents*. This statement was first made by Dalton, in 1801, and is called *Dalton's Law of Partial Pressures*.

Volume and Temperature: Charles's Law

Measurements of the change in volume of a gas with temperature, at constant pressure, were published by Charles in 1787 and by Gay-Lussac in 1802. Fig. 10.3 shows an apparatus which we may use for repeating their experiments. Air is trapped by mercury in the closed limb C of the tube AC; a scale engraved upon C enables us to measure the length of the air column, l . The tube is surrounded by a water-bath W, which we can heat by passing in steam. After making the temperature uniform by stirring, we level the mercury in the limbs A and C, by pouring mercury in at A, or running it off at B. The air

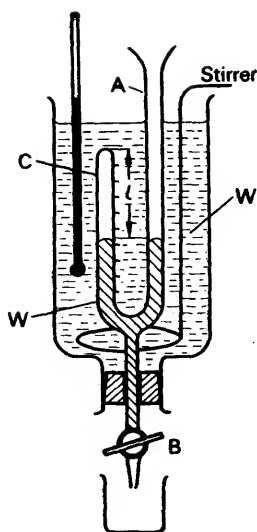


FIG. 10.3. Charles' law experiment.

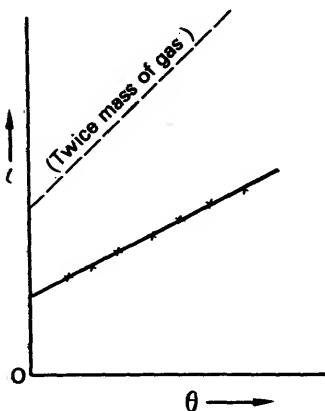


FIG. 10.4. Results of experiment.

in C is then always at atmospheric (constant) pressure. We measure the length l and plot it against the temperature, θ (Fig. 10.4).

If S is the cross-section of the tube, the volume of the trapped air is

$$V = lS.$$

The cross-section S , and the distance between the divisions on which we read l , both increase with the temperature θ . But their increases are very small compared with the expansion of the gas, and therefore we may say that the volume of the gas is proportional to the scale-reading of l . The graph then shows that the volume of the gas, at constant pressure, increases uniformly with its temperature. A similar result is obtained with twice the mass of gas, as indicated in Fig. 10.4.

Expansivity of Gas (Volume Coefficient)

We express the rate at which the volume of a gas increases with temperature by defining a quantity called its *expansivity at constant pressure*, α_p , or *volume coefficient*:

$$\alpha_p = \frac{\text{volume at } \theta^\circ\text{C} - \text{volume at } 0^\circ\text{C}}{\text{volume at } 0^\circ\text{C}} \times \frac{1}{\theta}$$

Thus, if V is the volume at $\theta^\circ\text{C}$, and V_0 the volume at 0°C , then

$$\alpha_p = \frac{V - V_0}{V_0 \theta},$$

whence

$$V - V_0 = V_0 \alpha_p \theta,$$

or

$$V = V_0(1 + \alpha_p \theta).$$

The expansivity α_p has the dimensions

$$\frac{[\text{volume}]}{[\text{volume}] \times [\text{temperature}]} = \frac{1}{[\text{temperature}]}$$

Its value is about $\frac{1}{273}$ when the temperature is measured in $^{\circ}\text{C}$, and we therefore say that

$$\alpha_p = \frac{1}{273} \text{ per deg C, or } \frac{1}{273} \text{ K}^{-1}.$$

Charles, and Gay-Lussac, found that α_p had the same value, $\frac{1}{273}$, for all gases. This observation is now called Charles's or Gay-Lussac's Law: *The volume of a given mass of any gas, at constant pressure, increases by $\frac{1}{273}$ of its value at 0°C , for every degree Centigrade rise in temperature.*

Absolute Temperature

Charles's Law shows that, if we plot the volume V of a given mass of any gas at constant pressure against its temperature θ , we shall get a straight line graph A as shown in Fig. 10.5. If we produce this line back-

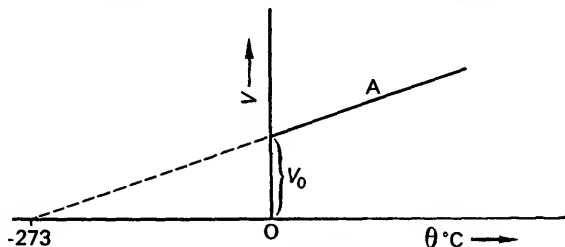


FIG. 10.5. Absolute zero.

wards, it will meet the temperature axis at -273°C . This temperature is called the *absolute zero*. If a gas is cooled, it liquefies before it reaches -273°C , and Charles's Law no longer holds; but that fact does not affect the form of the relationship between the volume and temperature at higher temperatures. We may express this relationship by writing

$$V \propto (273 + \theta).$$

The quantity $(273 + \theta)$ is called the *absolute temperature* of the gas, and is denoted by T . The idea of absolute temperature was developed by Lord Kelvin, and absolute temperatures are hence expressed in degrees Kelvin:

$$T \text{ K} = (273 + \theta)^{\circ}\text{C}.$$

From Charles's Law, we see that the volume of a given mass of gas at constant pressure is proportional to its absolute temperature, since

$$V \propto (273 + \theta) \propto T.$$

Thus if a given mass of gas has a volume V_1 at $\theta_1^{\circ}\text{C}$, and is heated at constant pressure to $\theta_2^{\circ}\text{C}$, its new volume is given by

$$\frac{V_1}{V_2} = \frac{273 + \theta_1}{273 + \theta_2} = \frac{T_1}{T_2}.$$

Pressure and Temperature

The effect of temperature on the pressure of a gas, at constant volume, was investigated by Amontons in 1702. His work was forgotten, however, and was re-discovered only after the work of Gay-Lussac and Charles on the effect of temperature on volume.

An apparatus for measuring the pressure of a constant volume of gas at various known temperatures is shown in Fig. 10.6 (a). The bulb B

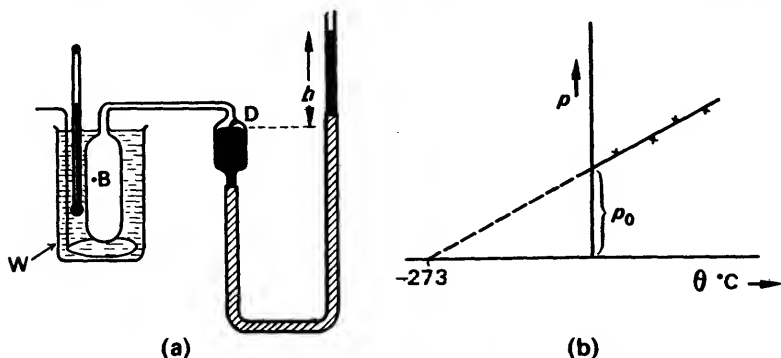


FIG. 10.6. Pressure and temperature.

contains air, which can be brought to any temperature θ by heating the water in the surrounding bath W. When the temperature is steady, the mercury in the closed limb of the tube is brought to a fixed level D, so that the volume of the air is fixed. The difference in level, h , of the mercury in the open and closed limbs is then added to the height of the barometer, H , to give the pressure p of the gas in cm of mercury. If p , $(h + H)$, is plotted against the temperature, the plot is a straight line (Fig. 10.6 (b)).

The coefficient of pressure increase at constant volume, α_v , known as the *pressure coefficient*, is given by

$$\alpha_v = \frac{p - p_0}{p_0 \theta},$$

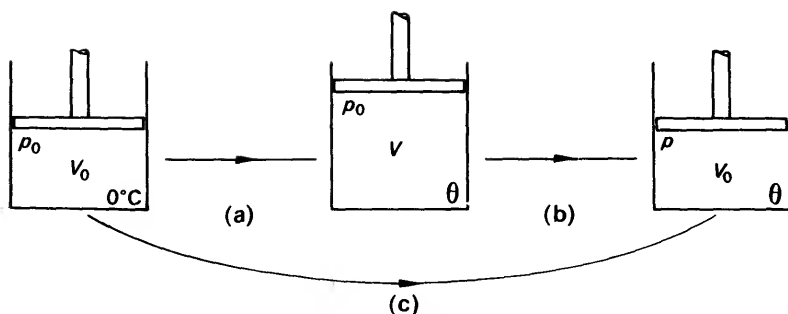
where p_0 is the pressure at 0°C . The coefficient α_v , which expresses the change of pressure with temperature, at constant volume, has practically the same value for all gases: $\frac{1}{273} \text{ K}^{-1}$. It is thus numerically equal to the expansivity, α_p . We may therefore say that, at constant volume, the pressure of a given mass of gas is proportional to its absolute temperature T , since

$$p \propto (273 + \theta).$$

$$\therefore \frac{p_1}{p_2} = \frac{273 + \theta_1}{273 + \theta_2} = \frac{T_1}{T_2}.$$

Equality of Pressure and Volume Coefficients

If a gas obeys Boyle's Law, its coefficient of pressure change at constant volume, α_v , and of volume change at constant pressure, α_p ,

FIG. 10.7. Showing that $\alpha_v = \alpha_p$.

must be equal. For let us suppose that a given mass of gas is warmed at constant pressure, p_0 , from 0°C to $\theta^\circ\text{C}$ (Fig. 10.7 (a)). Its volume expands from V_0 to V , where

$$V = V_0(1 + \alpha_p\theta).$$

Now let us suppose that it is compressed, at constant temperature, until its volume returns to V_0 (Fig. 10.7 (b)). Then its pressure rises to p , where

$$\begin{aligned} pV_0 &= p_0V \\ &= p_0V_0(1 + \alpha_p\theta) \end{aligned} \quad \text{or} \quad p = p_0(1 + \alpha_p\theta) \quad (3)$$

The condition of the gas is now the same as if it had been warmed at constant volume from 0°C to $\theta^\circ\text{C}$ (Fig. 10.7 (c)). Therefore

$$p = p_0(1 + \alpha_v\theta);$$

and, by equation (3), it follows that

$$\alpha_v = \alpha_p.$$

We shall see later that gases do not obey Boyle's law exactly, although at moderate pressures they do so very nearly. The difference between α_p and α_v provides a sensitive test for departures from Boyle's Law.

The Equation of State

Fig. 10.8 illustrates the argument by which we may find the general relationship between pressure, volume and temperature of a given mass of gas. This relationship is called the *equation of state*.

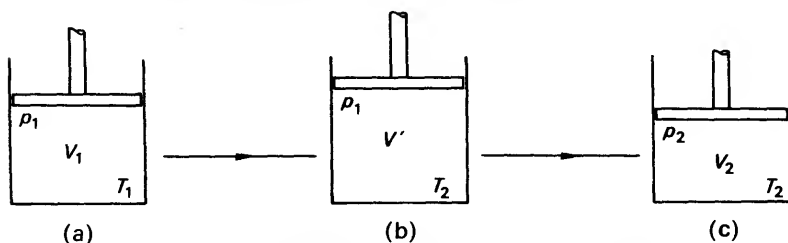


FIG. 10.8. Changing temperature and pressure of a gas.

At (a) we have the gas occupying a volume V_1 at a pressure p_1 , and an absolute temperature T_1 . We wish to calculate its volume V_2 at an absolute temperature T_2 and pressure p_2 , as at (c). We proceed via (b), raising the temperature of T_2 while keeping the pressure constant at p_1 . If V' is the volume of the gas at (b), then, by Charles's law:

$$\frac{V'}{V_1} = \frac{T_2}{T_1} \quad . \quad . \quad . \quad . \quad . \quad (4)$$

We proceed now to (c), by increasing the pressure to p_2 , while keeping the temperature constant at T_2 . By Boyle's law,

$$\frac{V_2}{V'} = \frac{p_1}{p_2} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Eliminating V' between equations (4) and (5), we find

$$\frac{V_2}{V_1} = \frac{T_2}{T_1} \cdot \frac{p_1}{p_2}$$

or

$$\frac{p_2 V_2}{T_2} = \frac{p_1 V_1}{T_1}.$$

In general therefore,

$$\frac{pV}{T} = R \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where R is a constant. This equation is often given in the form

$$pV = RT \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Equation (7) is the *equation of state* for a perfect gas. The value of the constant R depends on the nature of the gas—air, hydrogen, etc.—and on the mass of the gas concerned. If we consider *unit mass* of a gas, we can denote its volume by V and write

$$pV = RT \quad . \quad . \quad . \quad . \quad . \quad (8)$$

R is called the gas-constant for unit mass of the gas. If ρ is the density of the gas, at absolute temperature T and pressure p , then

$$\rho = \frac{1}{V},$$

and equation (8) becomes $\frac{p}{\rho} = RT \quad . \quad . \quad . \quad . \quad . \quad (9)$

The volume V of an arbitrary mass M of the gas, at absolute temperature T and pressure p , is

$$V = MV; \quad .$$

therefore, by (8) $pV = MRT; \quad . \quad . \quad . \quad . \quad . \quad (10)$

and, by (7), $R = MR.$

Magnitude of the Gas Constant

To calculate the constant R for a gas, we need to know the density of the gas at a given temperature and pressure. Very often, in dealing with gases, we specify the pressure not in newton per metre² (N m^{-2}) but simply in millimetres of mercury. 1 mm mercury pressure is called 1 *torr*. We do so because we are concerned only with relative values. A pressure of 760 mm of mercury, which is about the average pressure of the atmosphere, is sometimes called 'standard' or 'normal' pressure. A temperature of 0°C , or 273 K, is likewise called standard or normal temperature. The conditions 273 K and 760 mm pressure are together called standard temperature and pressure (s.t.p.). A pressure of 760 mm mercury is given, in newton per metre², by

$$p = g\rho H$$

$$= 9.8 \times 13600 \times 0.76 = 1.013 \times 10^5 \text{ N m}^{-2},$$

since ρ = mercury density = 13600 kg m^{-3} ; $g = 9.8 \text{ m s}^{-2}$; $H = 0.76 \text{ m}$.

At s.t.p. the density of hydrogen is about 0.09 g/litre, or 0.09 kg m^{-3} . The gas-constant for unit mass of hydrogen, from (9), is therefore

$$R = \frac{p}{\rho T} = \frac{1.013 \times 10^5}{0.09 \times 273}$$

$$= 4.16 \times 10^3, \text{ in the appropriate units.}$$

We will now discuss the units in which R is expressed.

The Gas-constant Units: Work done in Expansion

The gas-constant R for an *arbitrary mass* of gas is defined by the equation

$$pV = RT$$

or

$$R = \frac{pV}{T}.$$

Its unit is therefore that of

$$\frac{\text{pressure} \times \text{volume}}{\text{temperature}}.$$

In SI units, the pressure is in N m^{-2} , the volume in m^3 and the temperature in K. If we are given values of p , V and T and work out the value of R , we express it in the corresponding units. The constant per unit mass, R , for 1 kg has the unit of

$$\frac{\text{pressure} \times \text{volume}}{\text{temperature} \times \text{mass}} = \frac{\text{N m}^{-2} \times \text{m}^3}{\text{K} \times \text{kg}} = \text{N m kg}^{-1} \text{K}^{-1}$$

$$= \text{J kg}^{-1} \text{K}^{-1}.$$

since $1 \text{ newton} \times 1 \text{ metre} = 1 \text{ joule}$. The gas constant may thus be expressed in the same units as specific heat capacity (p. 199).

The gas constant depends on the mass of gas. For 1 kg, the unit is

$\text{J kg}^{-1} \text{K}^{-1}$. For 1 mole, the unit is $\text{J mol}^{-1} \text{K}^{-1}$; for 1 kmol, the unit is $\text{J kmol}^{-1} \text{K}^{-1}$.

Work Done. The product of pressure and volume has the dimensions of *work*. To see this, let us imagine some gas, at a pressure p , in a cylinder fitted with a piston (Fig. 10.9).

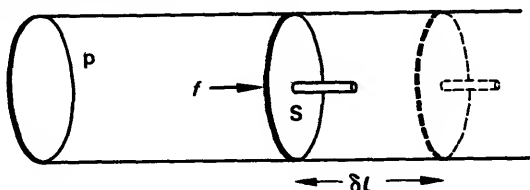


FIG. 10.9. Work done in expansion.

If the piston has an area S , the force on it is

$$f = pS.$$

If we allow the piston to move outwards a distance δl , the gas will expand, and its pressure will fall. But by making the distance very short, we can make the fall in pressure so small that we may consider the pressure constant. The force f is then constant, and the work done is

$$\delta W = f \cdot \delta l = pS \cdot \delta l.$$

The product $S \cdot \delta l$ is the increase in volume, δV , of the gas, so that

$$\delta W = p \cdot \delta V \quad (11)$$

The product of pressure and volume, in general, therefore represents work. If the pressure p is in newton m^{-2} , and the area S is in m^2 , the force f is in newtons. And if the movement δl is in m , the work $f \cdot \delta l$ is in newton \times metre or *joule* (J). The increase of volume, δV , is in m^3 . Thus the product of pressure in N m^{-2} , and volume in m^3 , represents work in joules.

Consequently, if we express the pressure of a gas in newton m^{-2} and its volume in m^3 , the gas-constant R , $= pV/T$, is in joule (J) per degree. The constant for 1 kg, R , is in joule per kg per degree. The value of R for hydrogen, which we calculated on p. 228, is

$$R = 4.16 \times 10^3 \text{ J kg}^{-1} \text{K}^{-1} = 4.16 \text{ kJ kg}^{-1} \text{K}^{-1}.$$

Avogadro's Hypothesis: Molecular Weight

Amedeo Avogadro, with one simple-looking idea, illuminated chemistry as Newton illuminated mechanics. In 1811 he suggested that chemically active gases, such as oxygen, existed not as single atoms, but as pairs: he proposed to distinguish between an atom, O, and a molecule, O_2 . Ampere proposed the same distinction, independently, in 1814. Avogadro also put forward another idea, now called *Avogadro's hypothesis*: that equal volumes of all gases, at the same temperature and pressure, contained equal numbers of molecules. The number of molecules in 1 cm^3 of gas at s.t.p. is called Loschmidt's number; it is 2.69×10^{19} .

Avogadro's hypothesis became accepted in the middle of the nineteenth century. Because molecules could not be observed, their masses could not be measured directly; but they could be compared, by chemical methods. The molecular mass of a substance, μ , was at first defined as the ratio of the mass of its molecule, m , to the mass of a hydrogen atom. Later, for the convenience of chemists, it was defined as the ratio of the molecular mass to the mass of an imaginary atom, this atom having $\frac{1}{12}$ th the mass of a carbon atom ^{12}C :

$$\mu = \frac{\text{mass of molecule}}{\frac{1}{12} \text{ mass of C-atom}} = \frac{m}{\frac{1}{12}m_{\text{C}}} = \frac{12m}{m_{\text{C}}}.$$

On this scale, the mass of a hydrogen atom is 1.008 times the mass of the imaginary atom. And since the hydrogen molecule contains two atoms, its molar mass is

$$\mu_{\text{H}_2} = 2.016.$$

The unit of molecular mass, $m_{\text{C}}/12$, is also the unit of atomic mass; its value is 1.66×10^{-24} g.

The Mole: Molar Gas-constant

The amount of a substance which contains as many elementary units as there are atoms in 0.012 kg (12 g) of carbon-12 is called a *mole*, symbol 'mol'. The number of molecules in a mole, N_{A} , is given, if m is the mass of a molecule in grammes, by

$$\mu = N_{\text{A}}m,$$

whence

$$\begin{aligned} N_{\text{A}} &= \frac{\mu}{m} = \frac{12m/m_{\text{C}}}{m} \\ &= \frac{12}{m_{\text{C}}}. \end{aligned}$$

The number of molecules per mole is thus the same for all substances. It is called Avogadro's constant, and is equal to $6.02 \times 10^{23} \text{ mol}^{-1}$.

From Avogadro's hypothesis, it follows that the mole of all gases, at the same temperature and pressure, occupy equal volumes. Experiment confirms this; at s.t.p. 1 mole of any gas occupies 22.4 litres. Consequently, if we denote by V the volume of 1 mole, then the ratio $\frac{pV}{T}$ is the same for all gases. We call it the *molar gas constant*, R , and

$$R = \frac{pV}{T}.$$

$$\text{At s.t.p.} \quad V = 22.4 \text{ litres} = 22.4 \times 10^{-3} \text{ m}^3$$

$$p = 760 \text{ mm mercury} = 1.013 \times 10^5 \text{ N m}^{-2} \text{ (p. 228)}$$

$$T = 273 \text{ K.}$$

$$\therefore R = \frac{1.013 \times 10^5 \times 22.4 \times 10^{-3}}{273}$$

$$= 8.31 \text{ J mol}^{-1} \text{ K}^{-1}.$$

The value of R is the same for the moles of all substances. If μ g is the molar mass of a gas, the constant for 1 g is thus

$$R = \frac{R}{\mu} \quad (12)$$

KINETIC THEORY

The kinetic theory of matter, which regards all bodies as assemblies of particles in motion—either vibrating or flying about—is an old one. Lucretius described it in the first century A.D. and Gassendi and Hooke revived it in the seventeenth century. In 1738 D. Bernoulli applied it in detail to a gas, and from it deduced Boyle's law, which was already known from experiment. Another century passed, however, before the kinetic view of a gas was fully developed—mainly by Clausius (1822–88), Boltzmann (1844–1906), and Maxwell (1831–79).

In the kinetic theory of gases, we seek to explain the behaviour of gases by considering the motion of their molecules. In particular, we suppose that the pressure of a gas is due to the molecules bombarding the walls of its container. Whenever a molecule bounces off a wall, its momentum at right-angles to the wall is reversed; the force which it exerts on the wall is equal to the rate of change of its momentum. The average force exerted by the gas on the whole of its container is the average rate at which the momentum of its molecules is changed by collision with the walls.

To find the pressure of the gas we must find this force, and then divide it by the area of the walls. The following assumptions are made to simplify the calculation:

- (a) The attraction between the molecules is negligible.
- (b) The volume of the molecules is negligible compared with the volume occupied by the gas.
- (c) The molecules are like perfectly elastic spheres.
- (d) The duration of a collision is negligible compared with the time between collisions.

Calculation of Pressure

Consider for convenience a cube of side l containing N molecules of gas each of mass m . Fig. 10.10. A typical molecule will have a velocity c at any instant and this will have components of u , v , w respectively in the direction of the three perpendicular axes Ox , Oy , Oz as shown. Thus $c^2 = u^2 + v^2 + w^2$.

Consider the force exerted on the face X of the cube due to the component u . Just before impact, the momentum of the molecule due to u is mu . After impact, the momentum is $-mu$, since the momentum reverses. Thus

$$\text{momentum change on impact} = mu - (-mu) = 2mu.$$

The time taken for the molecule to move across the cube to the opposite face and back to X is $2l/u$. Hence

$$\text{momentum change per second} = \frac{\text{momentum change}}{\text{time}}$$

$$= \frac{2mu}{2l/u} = \frac{mu^2}{l}$$

$$\therefore \text{force on X} = \frac{mu^2}{l}$$

$$\therefore \text{pressure on X} = \frac{\text{force}}{\text{area}} = \frac{mu^2}{l \times l^2} = \frac{mu^2}{l^3} \quad (i)$$

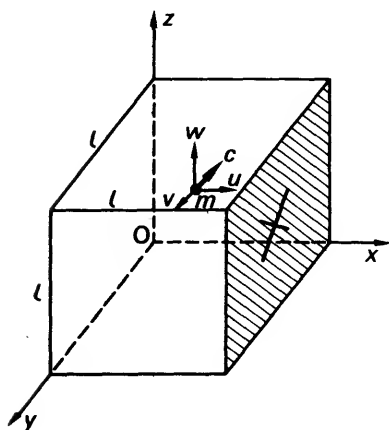


FIG. 10.10. Calculation of gas pressure.

We now take account of the N molecules in the cube. Each has a different velocity and hence a component of different magnitude in the direction Ox . If these are represented by $u_1, u_2, u_3, \dots, u_N$, it follows from (i) that the total pressure on X, p , is given by

$$p = \frac{mu_1^2}{l^3} + \frac{mu_2^2}{l^3} + \frac{mu_3^2}{l^3} + \dots + \frac{mu_N^2}{l^3}$$

$$= \frac{m}{l^3}(u_1^2 + u_2^2 + u_3^2 + \dots + u_N^2) \quad (ii)$$

Let the symbol $\overline{u^2}$ represent the average or mean value of all the squares of the components in the Ox direction, that is,

$$\overline{u^2} = \frac{u_1^2 + u_2^2 + u_3^2 + \dots + u_N^2}{N}.$$

Then
$$N\overline{u^2} = u_1^2 + u_2^2 + u_3^2 + \dots + u_N^2.$$

Hence, from (ii),

$$p = \frac{Nm\overline{u^2}}{l^3} \quad \text{. (iii)}$$

Now with a large number of molecules of varying speed in random motion, the mean square of the component speed in any one of the three axes is the *same*.

$$\therefore \overline{u^2} = \overline{v^2} = \overline{w^2}.$$

But, for each molecule, $c^2 = u^2 + v^2 + w^2$, so that the mean square $\overline{c^2}$ is given by $\overline{c^2} = \overline{u^2} + \overline{v^2} + \overline{w^2}$.

$$\therefore \overline{u^2} = \frac{1}{3}\overline{c^2}.$$

Hence, from (iii),

$$p = \frac{1}{3} \frac{Nm\overline{c^2}}{l^3}.$$

The number of molecules per unit volume, n , = N/l^3 . Thus we may write

$$p = \frac{1}{3} n m \overline{c^2} \quad \text{. (13)}$$

If n is in molecules per metre³, m is in kilogramme and c in metre per second, then the pressure p is in newton per metre² (N m^{-2}).

It should be carefully noted that the pressure p of the gas depends on the 'mean square' of the speed. This is because (a) the momentum change at a wall is proportional to u , as previously explained, and (b) the time interval before this change is repeated is *inversely*-proportional to u . Thus the rate of change of momentum is proportional to $u \div 1/u$ or to u^2 . Further, the mean-square speed is *not* equal to the square of the average speed. As an example, let us suppose that the speeds of six molecules are, 1, 2, 3, 4, 5, 6 units. Their mean speed is

$$\bar{c} = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5,$$

and its square is $(\bar{c})^2 = 3.5^2 = 12.25.$

Their mean square speed, however is

$$\overline{c^2} = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} = \frac{91}{6} = 15.2.$$

This differs by about 25 per cent from the square of the mean speed.

In our calculation, we assumed that molecules of a gas do not collide with other molecules as they move to-and-fro across the cube. If,

however, we assume that their collisions are perfectly elastic, both the kinetic energy and the momentum are conserved in them. The average momentum with which all the molecules strike the walls is then not changed by their collisions with one another; what one loses, another gains. The important effect of collisions between molecules is to distribute their individual speeds; on the average, the fast ones lose speed to the slow. We suppose, then, that different molecules have different speeds, and that the speeds of individual molecules vary with time, as they make collisions with one another; but we also suppose that the average speed of all the molecules is constant. These assumptions are justified by the fact that the kinetic theory leads to conclusions which agree with experiment.

Root-mean-square (R.M.S.) Speed

In equation (13) the factor nm is the product of the molecules per unit volume and the mass of one molecule. It is therefore the total mass of the gas per unit volume: its density ρ . Thus the equation gives

$$p = \frac{1}{3}\rho\overline{c^2} \quad . \quad . \quad . \quad . \quad (14)$$

or
$$\frac{p}{\rho} = \frac{1}{3}\overline{c^2} \quad . \quad . \quad . \quad . \quad (15)$$

If we substitute known values of p and ρ in equation (15), we can find $\overline{c^2}$. For hydrogen at s.t.p.,

$$\rho = 0.09 \text{ kg m}^{-3}.$$

The pressure in newton per m^2 , is $p = \rho g H$, where g = acceleration of gravity = 9.81 m s^{-2} , ρ = density of mercury = 13600 kg m^{-3} , H = barometer height = $760 \text{ mm} = 0.76 \text{ m}$.

$$\begin{aligned} \therefore \overline{c^2} &= \frac{3p}{\rho} = \frac{3 \times 9.81 \times 13600 \times 0.76}{9 \times 10^{-2}} \\ &= 3.37 \times 10^6 \text{ m}^2 \text{ s}^{-2}. \end{aligned}$$

The square root of $\overline{c^2}$ is called the *root-mean-square speed*; it is of the same magnitude as the average speed, but not quite equal to it. Its value is

$$\begin{aligned} \sqrt{\overline{c^2}} &= \sqrt{3.37 \times 10^6} = 1840 \text{ m s}^{-1} \text{ (approx.)} \\ &= 1.84 \text{ km s}^{-1} \end{aligned}$$

Molecular speeds were first calculated in this way by Joule in 1848; they turn out to have a magnitude which is high, but reasonable. The value is reasonable because it has the same order of magnitude as the speed of sound (1.30 km s^{-1} in hydrogen at 0°C); the speed of sound is the speed with which the molecules of a gas pass on a disturbance from one to another, and this we may expect to be of the same magnitude as the speeds of their natural motion.

Introduction of the Temperature

Let us consider a volume V of gas, containing N molecules. The number of molecules per unit volume is

$$n = \frac{N}{V},$$

and therefore the pressure of the gas, by equation (13) is

$$p = \frac{1}{3}nm\overline{c^2} = \frac{1}{3}\frac{N}{V}m\overline{c^2}$$

$$\therefore pV = \frac{1}{3}Nm\overline{c^2}. \quad (16)$$

Equation (16) reminds us of the equation combining Boyle's and Charles's laws:

$$pV = RT.$$

We can therefore make the kinetic theory consistent with the observed behaviour of a gas, if we write

$$\frac{1}{3}Nm\overline{c^2} = RT \quad (17)$$

Essentially, we are here assuming that the mean square speed of the molecules, $\overline{c^2}$, is proportional to the absolute temperature of the gas. This is a reasonable assumption, because we have learnt that heat is a form of energy; and the kinetic energy of a molecule, due to its random motion within its container, is proportional to the square of its speed. When we heat a gas, we expect to speed-up its molecules. See p. 237.

The kinetic energy of a molecule moving with a speed c is $\frac{1}{2}mc^2$; the average kinetic energy of translation of the random motion of the molecule of a gas is therefore $\frac{1}{2}m\overline{c^2}$. To relate this to the temperature, we put equation (17) into the form

$$RT = \frac{1}{3}Nm\overline{c^2} = \frac{2}{3}N(\frac{1}{2}m\overline{c^2}),$$

whence
$$\frac{1}{2}m\overline{c^2} = \frac{3}{2}\frac{R}{N}T \quad (18)$$

Thus, *the average kinetic energy of translation of a molecule is proportional to the absolute temperature of the gas.*

The ratio R/N in equation (18) is a universal constant. To see that it is, we have only to consider a mole. We have already seen that, for a mole, the gas constant R , and number of molecules N , are universal constants. If our arbitrary mass of gas is x moles, then $R = xR$, and $N = xN$; therefore

$$\frac{R}{N} = \frac{R}{N} = k.$$

The constant k , the gas constant per molecule, is also a universal constant; it is often called *Boltzmann's constant*. In terms of k equation (18) becomes

$$\frac{1}{2}m\overline{c^2} = \frac{3}{2}kT \quad (19)$$

Boltzmann's constant is usually given in joules per degree, since it relates energy to temperature:

$$k = \frac{\frac{1}{2}mc^2}{\frac{3}{2}T}.$$

Its value is $k = 1.38 \times 10^{-23} \text{ J K}^{-1}$.

Diffusion: Graham's Law

When a gas passes through a porous plug, a cotton-wool wad, for example, it is said to 'diffuse'. Diffusion differs from the flow of a gas through a wide tube, in that it is not a motion of the gas in bulk, but is a result of the motion of its individual molecules.

Fig. 10.11 shows an apparatus devised by Graham (1805–69) to compare the rates of diffusion of different gases. D is a glass tube, closed with a plug P of plaster of Paris. It is first filled with mercury, and inverted over mercury in a bowl. Hydrogen is then passed into it until the mercury levels are the same on each side; the hydrogen is then at atmospheric pressure. The volume of hydrogen, V_H , is proportional to the length of the tube above the mercury. The apparatus is now left; hydrogen diffuses out through P, and air diffuses in. Ultimately no hydrogen remains in the tube D. The tube is then adjusted until the level of mercury is again the same on each side, so that the air within it is at atmospheric pressure. The volume of air, V_A , is proportional to the new length of the tube above the mercury.

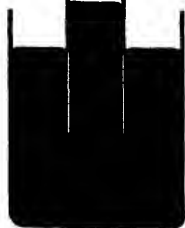


FIG. 10.11. Graham's apparatus for diffusion.

The volumes V_A and V_H are, respectively, the volumes of air and hydrogen which diffused through the plug in the same time. Therefore the rates of diffusion of the gases air and hydrogen are proportional to the volumes V_A and V_H :

$$\frac{\text{rate of diffusion of air}}{\text{rate of diffusion of hydrogen}} = \frac{V_A}{V_H}.$$

Graham found in his experiments that the volumes were inversely proportional to the square roots of the densities of the gases, ρ :

$$\frac{V_A}{V_H} = \sqrt{\frac{\rho_H}{\rho_A}}$$

thus
$$\frac{\text{rate of diffusion of air}}{\text{rate of diffusion of hydrogen}} = \sqrt{\frac{\rho_H}{\rho_A}}.$$

In general:
$$\text{rate of diffusion} \propto \frac{1}{\sqrt{\rho}};$$

and in words: *the rate of diffusion of a gas is inversely proportional to the square root of its density. This is Graham's Law.*

Graham's law of diffusion is readily explained by the kinetic theory. At the same Kelvin temperature T , the mean kinetic energies of the molecules of different gases are equal, since

$$\frac{1}{2}m\overline{c^2} = \frac{3}{2}kT$$

and k is a universal constant. Therefore, if the subscripts A and H denote air and hydrogen respectively,

$$\frac{1}{2}m_A\overline{c_A^2} = \frac{1}{2}m_H\overline{c_H^2},$$

whence

$$\frac{\overline{c_A^2}}{\overline{c_H^2}} = \frac{m_H}{m_A}.$$

At a given temperature and pressure, the density of a gas, ρ , is proportional to the mass of its molecule, m , since equal volumes contain equal numbers of molecules:

Therefore

$$\frac{m_H}{m_A} = \frac{\rho_H}{\rho_A},$$

whence

$$\begin{aligned} \frac{\overline{c_A^2}}{\overline{c_H^2}} &= \frac{\rho_H}{\rho_A} \\ \therefore \frac{\sqrt{\overline{c_A^2}}}{\sqrt{\overline{c_H^2}}} &= \frac{\sqrt{\rho_H}}{\sqrt{\rho_A}} \end{aligned} \quad (20)$$

The average speed of the molecules of a gas is roughly equal to—and strictly proportional to—the square root of its mean square speed. Equation (20) therefore shows that the average molecular speeds are inversely proportional to the square roots of the densities of the gases. And so it explains why the rates of diffusion—which depend on the molecular speeds—are also inversely proportional to the square roots of the densities.

Thermal Agitations and Internal Energy

The random motion of the molecules of a gas, whose kinetic energy depends upon the temperature, is often called the *thermal agitation* of the molecules. And the kinetic energy of the thermal agitation is called the *internal energy* of the gas. We must appreciate that this energy is quite independent of any motion of the gas in bulk: when a cylinder of oxygen is being carried by an express train, its kinetic energy as a whole is greater than when it is standing on the platform; but the random motion of the molecules within the cylinder is unchanged—and so is the temperature of the gas. The same is true of a liquid; in a water-churning experiment to convert mechanical energy into heat, baffles must be used to prevent the water from acquiring any mass-motion—all the work done must be converted into random motion, if it is to appear as heat. Likewise, the internal energy of a solid is the kinetic energy of its atoms' vibrations about their mean positions: throwing a lump of metal through the air does not raise its temperature, but hitting it with a hammer does.

The internal energy of a gas depends on the number of atoms in its molecule. A gas whose molecules consist of single atoms is said to be monatomic: for example, chemically inert gases and metallic vapours, Hg, Na, He, Ne, A. A gas

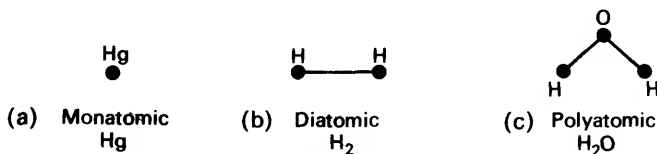


FIG. 10.12. Types of gas molecule.

with two atoms to the molecule is said to be diatomic: O₂, H₂, N₂, Cl₂, CO. And a gas with more than two atoms to the molecule is said to be polyatomic: H₂O, O₃, H₂S, CO₂, CH₄. The molecules of a monatomic gas we may regard as points, but those of a diatomic gas we must regard as 'dumb-bells', and those of a polyatomic gas as more complicated structures (Fig. 10.12). A molecule which extends appreciably in space—a diatomic or polyatomic molecule—has an

appreciable moment of inertia. It may therefore have kinetic energy of rotation, as well as of translation. A monatomic molecule, however, must have a much smaller moment of inertia than a diatomic or polyatomic; its kinetic energy of rotation can therefore be neglected.

Fig. 10.12A shows a monatomic molecule whose velocity c has been resolved into its components u , v , w along the x , y , z axes:

$$c^2 = u^2 + v^2 + w^2.$$

FIG. 10.12A. Components of velocity.

The x , y , z axes are called the molecules' degrees of freedom: they are three directions such that the motion of the molecule along any one is independent of its motion along the others.

If we average the speed c , and the components u , v , w , over all the molecules in a gas, we have

$$\overline{c^2} = \overline{u^2} + \overline{v^2} + \overline{w^2}.$$

And since the molecules do not pile up in any corner of the vessel containing the gas, their average velocities in all directions must be the same. We may therefore write

$$\overline{u^2} = \overline{v^2} = \overline{w^2},$$

$$\overline{c^2} = 3\overline{u^2} = 3\overline{v^2} = 3\overline{w^2},$$

$$\text{whence} \quad \overline{u^2} = \overline{v^2} = \overline{w^2} = \frac{1}{3}\overline{c^2}.$$

The average kinetic energy of a molecule of the gas is given by equation (19):

$$\frac{1}{2}m\overline{c^2} = \frac{3}{2}kT.$$

Therefore the average kinetic energy of a monatomic molecule, in each degree of freedom, is

$$\frac{1}{2}m\overline{u^2} = \frac{1}{2}m\overline{v^2} = \frac{1}{2}m\overline{w^2} = \frac{1}{2}kT.$$

Thus the molecule has kinetic energy $\frac{1}{2}kT$ per degree of freedom.

Rotational Energy

Let us now consider a diatomic or polyatomic gas. When two of its molecules collide, they will, in general, tend to rotate, as well as to deflect each other. In some collisions, energy will be transferred from the translations of the molecules

to their rotations; in others, from the rotations to the translations. We may assume, then, that the internal energy of the gas is shared between the rotations and translations of its molecules.

To discuss the kinetic energy of rotation, we must first extend the idea of degrees of freedom to it. A diatomic molecule can have kinetic energy of rotation about any axis at right-angles to its own. Its motion about any such axis can be resolved into motions about two such axes at right-angles to each other (Fig. 10.13 (a)). Motions about these axes are independent of each other, and a diatomic

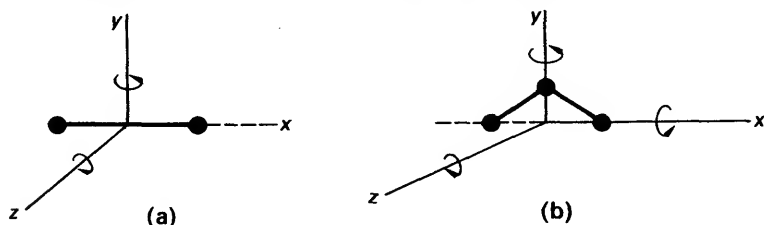


FIG. 10.13. Rotation of molecules.

molecule therefore has two degrees of rotational freedom. A polyatomic molecule, unless it happens to consist of molecules all in a straight line, has no axis about which its moment of inertia is negligible. It can therefore have kinetic energy of rotation about three mutually perpendicular axes (Fig. 10.13 (b)). It has three degrees of rotational freedom.

We have seen that the internal energy of a gas is shared between the translations and rotations of its molecules. Maxwell assumed that the average kinetic energy of a molecule, in each degree of freedom, *rotational as well as translational*, was $\frac{1}{2}kT$. This assumption is called the *principle of equipartition of energy*; experiment shows, as we shall find, that it is true at room temperature and above. At very low temperatures, when the gas is near liquefaction, it fails. At ordinary temperatures, then, we have:

$$\text{average K.E. of monatomic molecule} = \frac{3}{2}kT (\text{trans.});$$

$$\text{average K.E. of diatomic molecule} = \frac{3}{2}kT (\text{trans.}) + \frac{2}{2}kT (\text{rot.}) = \frac{5}{2}kT;$$

$$\text{average K.E. of polyatomic molecule} = \frac{3}{2}kT (\text{trans.}) + \frac{3}{2}kT (\text{rot.}) = \frac{6}{2}kT.$$

Internal Energy of any Gas

From the average kinetic energy of its molecules, we can find the internal energy of a mass M of a gas. The number of molecules in this mass is, if m is the mass of one molecule.

$$N = \frac{M}{m}.$$

Its internal energy, U , is the total kinetic energy of its molecules' random motions; thus

$$U = N \times \text{average K.E. of molecule.}$$

For a monatomic gas, therefore,

$$U = \frac{3}{2}NkT (\text{monatomic}).$$

The constant k is the gas-constant per molecule; the product Nk is therefore the gas-constant R for the mass M of the gas. Thus

$$U = \frac{3}{2}RT \text{ (monatomic).}$$

In particular, if R is the gas-constant per kg, the internal energy per kg is

$$U = \frac{3}{2}RT \text{ (monatomic)} \quad . \quad . \quad . \quad (21)$$

Similarly, for a diatomic gas,

$$U = \frac{5}{2}NkT = \frac{5}{2}RT \quad . \quad . \quad . \quad (22)$$

$$U = \frac{5}{2}RT$$

And for a polyatomic gas,

$$U = \frac{6}{2}NkT = \frac{6}{2}RT \quad . \quad . \quad . \quad (23)$$

$$U = \frac{6}{2}RT$$

Internal Energy and Volume

In our simple account of the kinetic theory of gases, we have implicitly assumed that the molecules of a gas do not attract one another. If they did, any molecule approaching the boundary of the gas would be pulled towards the body of it, as is a molecule of water approaching the surface (see Chapter 6, *Surface Tension*, p. 128). The attractions of the molecules would thus reduce the pressure of the gas.

Since the molecules of a substance are presumably the same whether it is liquid or gas, the molecules of a gas must attract one another somewhat. But except for brief instants when they collide, the molecules of a gas are much further apart than those of a liquid. In 1 cubic centimetre of gas at s.t.p. there are 2.69×10^{19} molecules, and in 1 cubic centimetre of water there are 3.33×10^{22} ; there are a thousand times as many molecules in the liquid, and so the molecules in the gas are ten times further apart. We may therefore expect that the mutual attraction of the molecules of a gas, for most purposes, can be neglected, as experiment, in fact, shows.

The experiment consists in allowing a gas to expand without doing external work; that is, to expand into a vacuum. Then, if the molecules attract one another, work is done against their attractions, as they move further apart. But if the molecular attractions are negligible, the work done is also negligible. If any work is done against the molecular attractions, it will be done at the expense of the molecular kinetic energies; as the molecules move apart, they will exert retarding forces on one another. Thus the internal energy of the gas, and therefore its temperature, will fall.

The expansion of a gas into a vacuum is called a 'free expansion'. If a gas does not cool when it makes a free expansion, then the mutual attractions of its molecules are negligible.

Joule's Experiments

Experiments on the free expansion of a gas were made in 1807 by Gay-Lussac; they showed no fall in temperature. Joule repeated these experiments with a better vacuum in 1845; he got the same negative result, and the greater accuracy of his experiments made them more trustworthy. Joule used two forms of apparatus, as shown in Fig. 10.14 (a) and (b). Each consisted of a cylinder of air, R, at 22 atmospheres, connected by a stop-cock S to an evacuated cylinder E. In the apparatus (a) both cylinders stood in the same tin can C, which contained $16\frac{1}{2}$ lb water. In (b) the cylinders stood in different cans, and the stop-cock in a third, also containing water. When the stop-cock was opened, gas expanded from R to E. With the apparatus (a) Joule found, after stirring the water, that its temperature was unchanged. The expanding gas had therefore neither liberated heat nor absorbed it. With the apparatus (b) Joule found that heat was absorbed from the water round R, and given to the water round S and E; the heat given out was equal to the heat

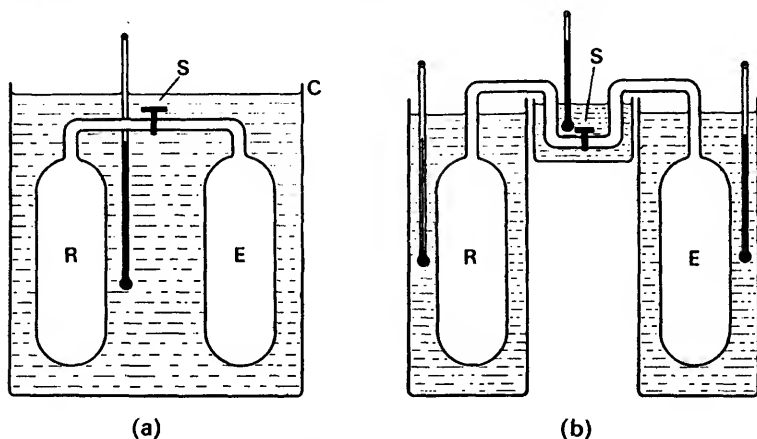


FIG. 10.14. Joule's experiment on internal energy.

taken in. The heat taken in represented the work done by the gas from R, expanding against the rising pressure of the gas in E and in the pipe beyond S. The heat given out represented the work done on the gas in E and S by the gas flowing in, against the rising pressure. The equality of the two showed that the total mass of gas neither gained nor lost energy in making its free expansion. Joule's experiments, therefore, showed that *the internal energy of a gas is independent of its volume*.

From Joule's results we may argue back to show that the mutual attractions of the molecules of the gas are negligible; in practice, however, it is the property of the bulk gas which is important—the fact that its internal energy does not depend on its volume.

Joule's experiments, though more reliable than Gay-Lussac's, were crude; with so much water, a small amount of heat would not produce a measurable temperature rise. Between 1852 and 1862, Joule worked with William Thomson, later Lord Kelvin, on more delicate experi-

ments. They found that most gases, in expanding from high pressure to low, do lose a little of their internal energy. The loss represents work done against the molecular attractions, which are therefore not quite negligible.

If the internal energy of a gas is independent of its volume, it is determined only by the temperature of the gas. The simple expression for the pressure, $p = \frac{1}{3}\rho c^2$, then holds; and the gas obeys Boyle's and Charles's laws. Its pressure coefficient, α_p , is equal to its volume coefficient, α_v . Such a gas is called an ideal, or perfect, gas. All gases, when far from liquefaction, behave for most practical purposes as though they were ideal.

Van der Waals' Equation

In deriving the ideal gas equation $pV = RT$ from the kinetic theory of gases, a number of assumptions were made. These are listed on p. 231. Van der Waals modified the ideal gas equation to take account that two of these assumptions may not be valid. Thus, as explained on p. 127, to which the student should refer:

- (1) *The volume of the molecules may not be negligible in relation to the volume V occupied by the gas.*
- (2) *The attractive forces between the molecules may not be negligible.*

Molecules have a particular diameter or volume because repulsive forces occur when they approach very closely and hence they can not be compressed indefinitely. The volume of the space inside a container occupied by the molecules is thus not V but $(V - b)$, where b is a factor depending on the actual volume of the molecules.

If the attractive forces between molecules are not negligible, the molecules approaching the container walls are attracted by the molecules behind them. This reduces the momentum of the approaching molecules and hence the pressure. The observed pressure p is thus less than the ideal gas pressure, where there are no molecular forces, by a pressure p' . Hence we write $(p + p')$ in place of p in the ideal gas equation. As explained on p. 127, the 'pressure defect' $p' \propto \rho^2$, where ρ is the density of the gas, or $p = a/V^2$, where a is a constant. Thus *van der Waals' equation* for real gases is:

$$\left(p + \frac{a}{V^2}\right)(V - b) = RT.$$

At high pressures, when the molecules are relatively numerous and close together, the volume factor b and pressure 'defect' a/V^2 both become important. Conversely, at low pressures, where the molecules are relatively few and far apart on the average, a gas behaves like an ideal gas and obeys the equation $pV = RT$.

Critical Phenomena

A graph of pressure p v. volume V at constant temperature is called an *isotherm*. Fig. 10.15 (i) shows a number of isotherms for an ideal gas, which obeys the perfect gas law $pV = RT$. Fig. 10.15 (ii) shows a

number of isotherms for a gas which obeys van der Waals' equation, $(p + a/V^2)(V - b) = RT$.

At high temperatures the isotherms are similar. As the temperature is lowered, however, the isotherms in Fig. 10.15 (ii) change in shape. One curve has a point of inflexion at C, which corresponds to the *critical point* of a real gas. The isotherms thus approximate to those obtained by Andrews in his experiments on actual gases such as carbon dioxide, described on p. 314.

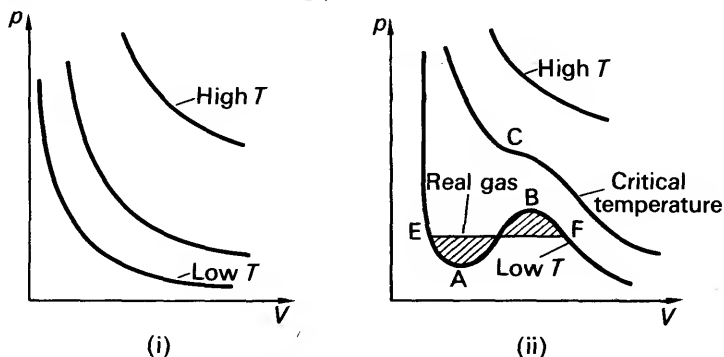


FIG. 10.15. Isotherms for ideal and van der Waals gases.

Below this temperature, however, isotherms such as EABF are obtained by using van der Waals' equation. These are unlike the isotherms obtained with real gases, because in the region AB the pressure increases with the volume, which is impossible. However, an actual isotherm in this region corresponds to a straight line EF, as shown. Here the liquid and vapour are in equilibrium (see p. 317) and the line EF is drawn to make the shaded areas above and below it equal. Thus van der Waals' equation roughly fits the isotherms of actual gases above the critical temperature but below the critical temperature it must be modified considerably. Many other gas equations have been suggested for real gases but quantitative agreement is generally poor.

SPECIFIC HEAT CAPACITIES

Specific Heat Capacities at Constant Volume and Constant Pressure

When we warm a gas, we may let it expand or not, as we please. If we do not let it expand—if we warm it in a closed vessel—then it does no external work, and all the heat we give it goes to increase its internal energy. *The heat required to warm unit mass of a gas through one degree, when its volume is kept constant, is called the specific heat capacity of the gas at constant volume.* It is denoted by c_v , and is generally expressed in $\text{J kg}^{-1} \text{K}^{-1}$.

If we allow a gas to expand as we warm it, then it does external work. The heat we give the gas appears partly as an increase to its internal energy—and hence its temperature—and partly as the heat equivalent of the work done. The work done depends on the increase in volume of the gas, which in turn depends on the way in which we allow the

gas to expand. We can get an important theoretical result by supposing that the pressure is constant, and defining the corresponding specific heat capacity. *The specific heat capacity of a gas at constant pressure is the heat required to warm unit mass of it by one degree, when its pressure is kept constant.* It is denoted by c_p , and is expressed in the same units as c_v .

Specific Heat Capacities: their Difference

Any number of heat capacities can be defined for a gas, according to the mass and the conditions imposed upon its pressure and volume. For unit mass, 1 kg or 1 g, of a gas, the heat capacities at constant pressure c_p , and at constant volume c_v , are the *specific heat capacities*.

Fig. 10.16 shows how we can find a relationship between the specific heat capacities of a gas. We first consider 1 kg of the gas warmed through

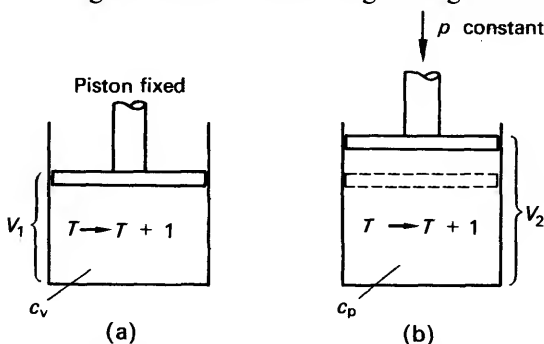


FIG. 10.16. Specific heat capacity at constant volume and pressure.

1°C at constant volume, (a). The heat required is c_v joules, and goes wholly to increase the internal energy.

We next consider 1 kg warmed through 1°C at constant pressure, (b). It expands from V_1 to V_2 , and does an amount of external work given by

$$W = p(V_2 - V_1) \quad (\text{equation (11), p. 229}).$$

The work W is in joules if p is in newton m^{-2} , and the volumes in m^3 . Thus the amount of heat in joules required for this work is

$$W = p(V_2 - V_1).$$

Further, since the temperature rise of the gas is 1°C, and the internal energy of the gas is independent of volume, the rise in internal energy is c_v , the specific heat at constant volume. Hence, from $\delta Q = \delta U + p \cdot \delta V$, the total amount of heat required to warm the gas at constant pressure is therefore

$$c_p = c_v + p(V_2 - V_1) \quad (24)$$

We can simplify the last term of this expression by using the equation of state for unit mass:

$$pV = RT,$$

where T is the absolute temperature of the gas, and R is the gas-

This was shown on p. 240.

Hence
$$c_v = \frac{5}{2}R$$

and
$$c_p = c_v + R = \frac{7}{2}R$$

Hence
$$\gamma = \frac{c_p}{c_v} = \frac{7}{5} = 1.40.$$

And for a polyatomic molecule,

$$U = \frac{6}{2}RT, \quad \text{. (iii)}$$

$$c_v = \frac{6}{2}R$$

and
$$\gamma = \frac{c_p}{c_v} = \frac{8}{6} = 1.33.$$

In general, if the molecules of a gas have f degrees of freedom, the average kinetic energy of a molecule is $f \times \frac{1}{2}kT$ (p. 238).

$$\therefore U = \frac{f}{2}RT,$$

$$c_v = \frac{f}{2}R,$$

$$c_p = c_v + R = \left(\frac{f}{2} + 1\right)R,$$

and
$$\gamma = \frac{c_p}{c_v} = \frac{\frac{f}{2} + 1}{\frac{f}{2}} = 1 + \frac{2}{f} \quad \text{. (iv)}$$

The ratio of the specific heat capacities of a gas thus gives us a measure of the number of atoms in its molecule, at least when that number is less than three. This ratio is fairly easy to measure, as we shall see later in this chapter. The poor agreement between the observed and theoretical values of γ for some of the polyatomic gases shows that, in its application to such gases, the theory is oversimple.

Measurement of c_v .

Fig. 10.17 shows an apparatus for measuring the specific heat capacity of a gas at constant volume, called a *differential steam calorimeter*. The calorimeter consists of two copper globes, A and B, as nearly identical as they can be made. They hang from the beam of a balance, and are surrounded by a chest C into which steam can be admitted at D. The sphere B is evacuated and A is filled with the gas whose specific heat is required. By filling A to a high pressure, the mass of gas can be made great enough to be accurately measurable on the balance. Let its value be M . Steam is now admitted to the chest, and condenses on both globes until they reach the temperature of the

steam. This will generally be about 100°C , but we shall denote it by θ_1 . The balance measures the excess steam condensed on A, over that condensed on B; let its mass be M_s . If the globes are identical, their heat capacities are equal, and the masses of steam required to warm the globes alone are equal. The excess steam condensed on A is then the mass required to warm the gas within it. Therefore if θ_R is the room temperature, and l the specific latent heat of evaporation of water, the specific heat capacity c_v for the gas is given by

$$Mc_v(\theta_1 - \theta_R) = M_s l;$$

whence

$$c_v = \frac{M_s l}{M(\theta_1 - \theta_R)} \quad (28)$$

The calorimeter is called 'differential' for the reason that it measures the difference in mass of the steam on the two globes. This is an important feature of it, because the heat capacities of the globes may be much greater than that of the gas. If a single globe were used, two measurements would have to be made, one with and one without the

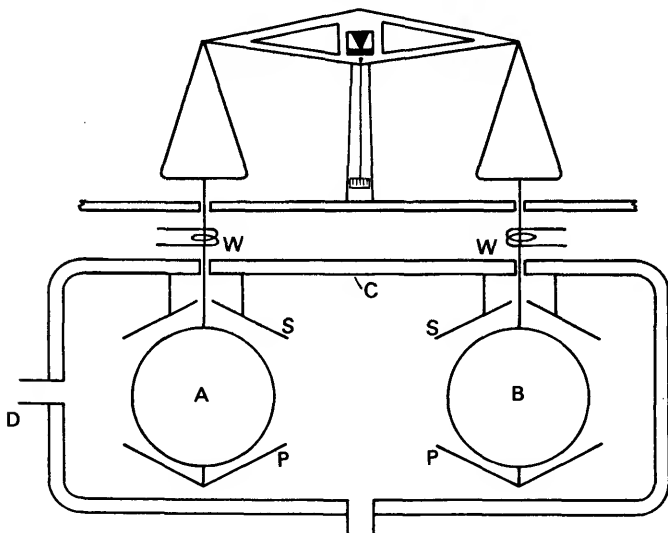


FIG. 10.17. Joly's differential steam calorimeter.

gas. The mass of steam condensed by the gas would then appear as the difference of two nearly equal masses, and could not be determined accurately. In practice the globes are not identical, and a control experiment with both evacuated is made to find the difference in mass of the steam condensed on them. This appears as a small correction to M_s in equation (28).

A small correction has also to be made to the result of the experiment, because the volume of the gas is not quite constant: the globe A expands when it is warmed.

The figure shows a few of the practical refinements of the apparatus. S, S are shields to prevent drops of moisture, condensed on the roof of

the chest, from falling on to the globes. P, P are pans to catch any drops which, having condensed on A or B, might fall off. W, W are platinum wires heated by an electric current, which prevent drops forming in the holes through which the suspension wires pass out of the chest.

Measurement of c_p .

The method of mixtures was used to determine the specific heat capacity of a gas at constant pressure by Regnault. Regnault was one of the greatest experimenters of the nineteenth century—the reader who sees pictures of his apparatus in other books should remember that he worked before Bunsen had invented his famous burner—but his method for c_p is now outmoded. We shall describe here only a continuous flow method, similar to Callendar and Barnes' for the specific heat capacity of water. It is due to Swann (1909).

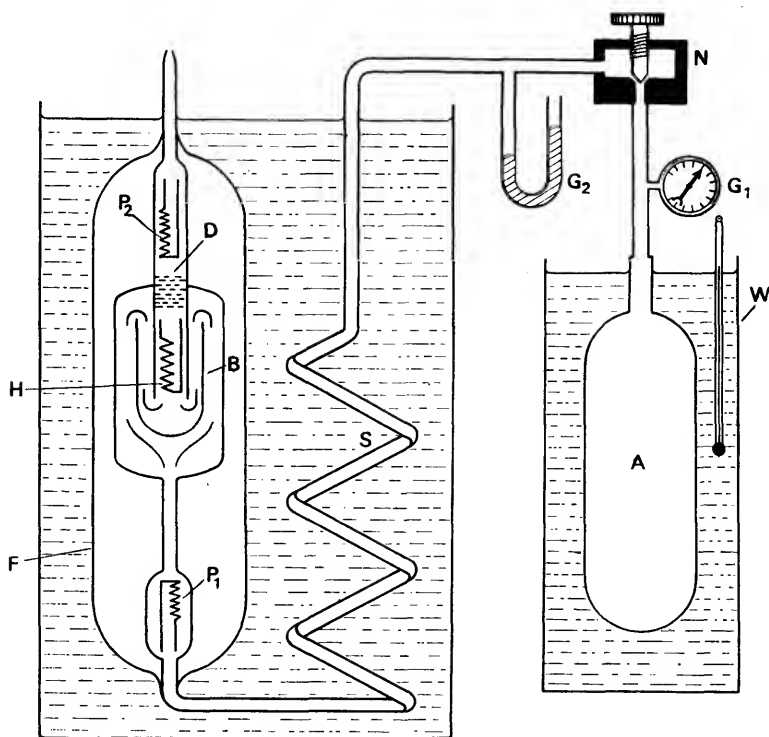


FIG. 10.18. Constant flow calorimeter for c_p .

Gas from a cylinder, A in Fig. 10.18, flows out through a needle valve N, which reduces its pressure to a little above atmospheric. If the pressure in the cylinder is high, it will fall slowly during an experiment, and the pressure of the emerging gas will be almost constant. Manometers G_1 G_2 indicate the pressure of the gas in the cylinder, and of the gas emerging. The gas passes through a coiled tube S, in a water bath,

which brings it to a uniform temperature. It then flows past a platinum resistance thermometer, P_1 , which measures the temperature, $\theta_1^\circ\text{C}$. From there it goes to a heating coil H, past a baffle B which enables it to receive any heat that escapes from the neighbourhood of the coil. Beyond the coil it passes through copper gauze D, which mixes the stream of gas and so brings it to a uniform temperature. This temperature, θ_2 , is measured by the platinum resistance thermometer P_2 . A vacuum jacket F makes the heat losses very small.

If M is the mass of gas flowing through the apparatus in t seconds, then the heat received by it is $Mc_p(\theta_2 - \theta_1)$. If the heat losses are negligible, the heat supplied by the coil in t seconds is IVt joule, where I is the current through it in ampere and V the potential difference across it in volt. Then

$$IVt = Mc_p(\theta_2 - \theta_1).$$

The mass of gas, M , is found from the fall of pressure in the cylinder. If v is the volume of the cylinder, and ρ_1 the density of the gas at the initial pressure p_1 , then the mass initially in the cylinder is

$$M_1 = \rho_1 v.$$

And if, after t seconds, the pressure has fallen to p_2 , and the density to ρ_2 , the mass remaining in the cylinder is

$$M_2 = \rho_2 v.$$

The mass of gas which has escaped in then

$$M = M_1 - M_2 = (\rho_1 - \rho_2)v.$$

The densities ρ_1 and ρ_2 can be readily calculated from the density ρ_0 at s.t.p.: if θ_3 is the temperature of the cylinder, and the pressures are measured in mm mercury, then

$$\frac{p_1}{\rho_1(273 + \theta_3)} = \frac{p_2}{\rho_2(273 + \theta_3)} = \frac{760}{273\rho_0}.$$

The cylinder temperature θ_3 is kept constant by the water bath W.

CHANGES OF PRESSURE, VOLUME AND TEMPERATURE

In, for example, a steam engine or motor, gases expand and are compressed, cool and are heated, in ways more complicated than those which we have already described. We shall now consider some of these ways.

Isothermal Changes

We have seen that the pressure p , and volume V of a given mass of gas are related by the equation

$$pV = RT,$$

where T is the absolute temperature of the gas, and R is a constant.

If the temperature is constant the curve of pressure against volume is a rectangular hyperbola,

$$pV = \text{constant},$$

representing Boyle's law. Such a curve is called an *isothermal* for the given mass of the given gas, at the temperature T . Fig. 10.19 shows a family of isotherms, for 1 g of air at different temperatures. When a gas expands, or is compressed, at *constant temperature*, its pressure and volume vary along the appropriate isothermal, and the gas is said to undergo an *isothermal compression or expansion*.

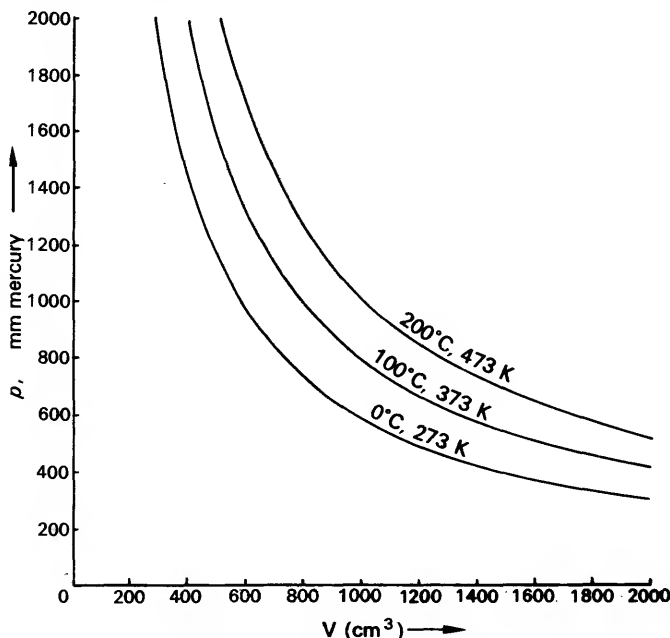


FIG. 10.19. Isotherms for 1 g air.

When a gas expands, it does work—for example, in driving a piston (Fig. 10.9, p. 229). The molecules of the gas bombard the piston, and if the piston moves they give up some of their kinetic energy to it; when a molecule bounces off a *moving* piston, it does so with a velocity less in magnitude than that with which it struck. The change in velocity is small, because the piston moves much more slowly than the molecule; but there are many molecules striking the piston at any instant, and their total loss of kinetic energy is equal to the work done in driving the piston forward.

The work done by a gas in expanding, therefore, is done at the expense of its internal energy. The temperature of the gas will consequently fall during expansion, unless heat is supplied to it. For an isothermal expansion, the gas must be held in a thin-walled, highly conducting vessel, surrounded by a constant temperature bath. And the expansion must take place slowly, so that heat can pass into the gas to maintain its temperature at every instant during the expansion.

External Work done in Expansion

The heat taken in when a gas expands isothermally is the heat equivalent of the mechanical work done. If the volume of the gas increases by a small amount δV , at the pressure p , then the work done is

$$\delta W = p\delta V \text{ (equation (11), p. 229).}$$

In an expansion from V_1 to V_2 , therefore, the work done is

$$W = \int dW = \int_{V_1}^{V_2} p dV.$$

By the gas equation, $p = \frac{RT}{V}$,

whence
$$W = \int_{V_1}^{V_2} p dV = \int_{V_1}^{V_2} RT \frac{dV}{V}$$

or
$$W = RT \log_e \left(\frac{V_2}{V_1} \right).$$

The heat required, Q , is therefore

$$Q = W = RT \log_e \left(\frac{V_2}{V_1} \right),$$

where W is in joules if R is in $\text{J kg}^{-1} \text{K}^{-1}$.

Now let us consider an isothermal compression. When a gas is compressed, work is done on it by the compressing agent. To keep its temperature constant, therefore, heat must be withdrawn from the gas, to prevent the work done from increasing its internal energy. The gas must again be held in a thin well-conducting vessel, surrounded by a constant-temperature bath; and it must be compressed slowly.

The conditions for an isothermal compression or expansion of a gas are difficult to realize; heat cannot flow through the walls of the vessel unless there is at least a small difference of temperature across them, and therefore the temperature of the gas is bound to rise a little in compression, or to fall a little in expansion.

Reversible isothermal change

Suppose a gas expands isothermally from p_1, V_1, T to p_2, V_2, T . If the change can be reversed so that the state of the gas is returned from p_2, V_2, T to p_1, V_1, T through exactly the same values of pressure and volume at every stage, then the isothermal change is said to be *reversible*. A reversible isothermal change is an ideal one. It requires conditions such as a light frictionless piston, so that the pressure inside and outside the gas can always be equalised and no work is done against friction; very slow expansion, so that no eddies are produced in the gas to dissipate the energy; and a constant temperature reservoir with very thin good-conducting walls, as we have seen. In a reversible isothermal change, $pV = \text{constant} = RT$.

Equation for Reversible Adiabatic change

Let us now consider a change of volume in which the conditions are at the opposite extreme from isothermal; no heat is allowed to enter or leave the gas.

An expansion or contraction in which *no heat* enters or leaves the gas is called an *adiabatic expansion or contraction*. In an adiabatic expansion, the external work is done wholly at the expense of the internal energy of the gas, and the gas therefore cools. In an adiabatic compression, all the work done on the gas by the compressing agent appears as an increase in its internal energy and therefore as a rise in its temperature. We have already discussed a reversible isothermal change. A *reversible adiabatic change* is an adiabatic change which can be exactly reversed in the sense explained on p. 251. As noted there, a reversible change is an ideal case.

The curve relating pressure and volume for a given mass of a given gas for adiabatic changes is called an 'adiabatic'. In Fig. 10.20, the heavy curve is an adiabatic for 1 g of air; it is steeper, at any point, than the isothermal through that point. The curve AB is the isothermal for the temperature $T_0 = 373$ K, which cuts the adiabatic at the point $p_0 V_0$. If the gas is adiabatically compressed from V_0 to V_1 , its

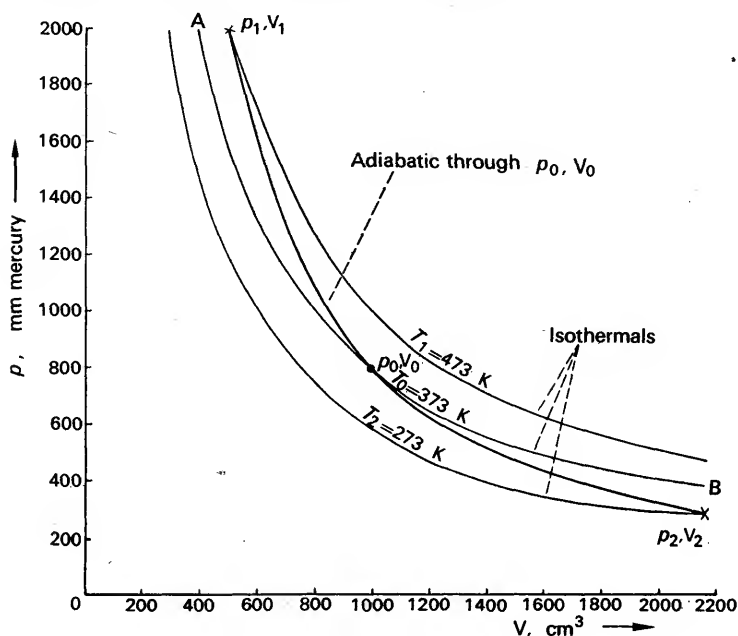


FIG. 10.20. Relationship between adiabatic and isothermals.

temperature rises to some value T_1 . Its representative point p_1, V_1 now lies on the isothermal for T_1 , since $p_1 V_1 = RT_1$. Similarly, if the gas is expanded adiabatically to V_2 , it cools to T_2 and its representative point p_2, V_2 lies on the isothermal for T_2 . Thus the adiabatic through

any point—such as p_0 , V_0 —is steeper than the isothermal. We will find its equation shortly.

The condition for an adiabatic change is that no heat must enter or leave the gas. The gas must therefore be held in a thick-walled, badly conducting vessel; and the change of volume must take place rapidly, to give as little time as possible for heat to escape. However, in a rapid compression, for example, eddies may be formed, so that some of the work done appears as kinetic energy of the gas in bulk, instead of as random kinetic energy of its molecules. All the work done then does not go to increase the internal energy of the gas, and the temperature rise is less than in a truly adiabatic compression. If the compression is made slowly, then more heat leaks out; since no vessel has perfectly insulating walls:

Perfectly adiabatic changes are therefore impossible; and so, we have seen, are perfectly isothermal ones. Any practical expansion or compression of a gas must lie between isothermal and adiabatic. It may lie anywhere between them, but if it approximates to isothermal, the curve representing it will always be a little steeper than the ideal (Fig. 10.21); if it approximates to adiabatic, the curve representing it will never be quite as steep as the ideal.

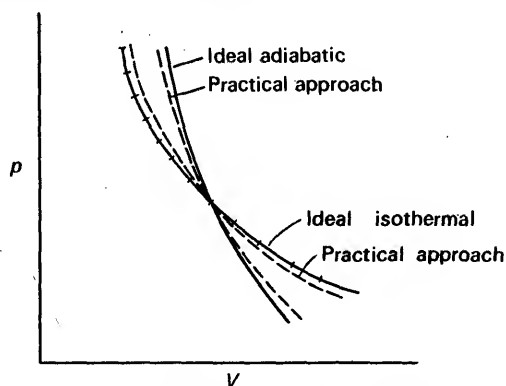


FIG. 10.21. Ideal and real p - V curves for a gas.

Equation of Reversible Adiabatic

Before considering adiabatic changes in particular, let us first consider a change of volume and temperature which takes place in an arbitrary manner. For simplicity, we consider unit mass of the gas, and we suppose that its volume expands from V to $V + \delta V$, and that an amount of heat δQ is supplied to it. In general, the internal energy of the gas will increase by an amount δU . And the gas will do an amount of external work equal to $p\delta V$, where p is its pressure. The heat supplied is equal to the increase in internal energy, plus the external work done:

$$\delta Q = \delta U + p\delta V \quad (29)$$

The increase in internal energy represents a temperature rise, δT . We have seen already that the internal energy is independent of the volume,

and is related to the temperature by the specific heat capacity at constant volume, c_v (p. 244). Therefore

$$\delta U = c_v \delta T.$$

Equation (29) becomes

$$\delta Q = c_v \delta T + p \delta V \quad (30)$$

Equation (30) is the fundamental equation for any change in the state of unit mass of a gas.

For a reversible isothermal change, $\delta T = 0$, and $\delta Q = p \delta V$.

For a reversible adiabatic change, $\delta Q = 0$ and therefore

$$c_v \delta T + p \delta V = 0 \quad (31)$$

To eliminate δT we use the general equation, relating pressure, volume and temperature:

$$pV = RT,$$

where R is the gas constant for unit mass. Since both pressure and volume may change, when we differentiate this to find δT we must write

$$p \delta V + V \delta p = R \delta T,$$

whence

$$\delta T = \frac{p \delta V + V \delta p}{R}.$$

Therefore, by equation (31),

$$c_v \frac{p \delta V + V \delta p}{R} + p \delta V = 0$$

or

$$c_v(p \delta V + V \delta p) + R p \delta V = 0.$$

Now we have seen, on p. 245, that

$$R = c_p - c_v;$$

therefore

$$c_v(p \delta V + V \delta p) + (c_p - c_v)p \delta V = 0.$$

Hence

$$c_v V \delta p + c_p p \delta V = 0$$

or

$$V \delta p + \frac{c_p}{c_v} p \delta V = 0$$

or

$$V \delta p + \gamma p \delta V = 0 \quad \left(\text{where } \gamma = \frac{c_p}{c_v} \right).$$

Therefore

$$\frac{\delta p}{p} + \gamma \frac{\delta V}{V} = 0.$$

Integrating, we find

$$\int \frac{dp}{p} + \gamma \int \frac{dV}{V} = 0$$

or

$$\log_e p + \gamma \log_e V = A,$$

where A is a constant.

Therefore,

$$pV^\gamma = C,$$

where C is also a constant. This is the equation of a reversible adiabatic; the value of C can be found from the initial pressure and volume of the gas.

If we have a mass M of the gas, its volume at any temperature and pressure is

$$V = MV,$$

where V is the volume of unit mass at the same temperature and pressure. Therefore for any mass of gas, the equation of an adiabatic change is

$$pV^\gamma = \text{constant} \quad (32)$$

Equation for Temperature Change in an Adiabatic.

If we wish to introduce the temperature, T , into equation (32), we use the general gas equation

$$pV = RT.$$

Thus

$$p = \frac{RT}{V}$$

and

$$pV^\gamma = \frac{RT}{V} \cdot V^\gamma = RTV^{\gamma-1}.$$

Thus equation (32) becomes

$$RTV^{\gamma-1} = \text{constant},$$

and since R is a constant for a given mass of gas, the equation for an adiabatic temperature change becomes

$$TV^{\gamma-1} = \text{constant}.$$

Measurement of γ .

In books on Sound, it is shown that sound waves are propagated through a gas by rapid compressions and rarefactions; these changes in pressure and volume are adiabatic. In consequence, the velocity of sound in a gas depends upon the ratio of the specific heat capacities of the gas, γ ; and the value of γ can be found from measurements of the velocity of sound in the gas. This is the most convenient way of measuring γ .

A direct measurement of γ can be made by the method of Clément and Désormes (1819). A large vessel—such as a carboy—contains the gas, which in a teaching experiment is usually air (Fig. 10.22). The

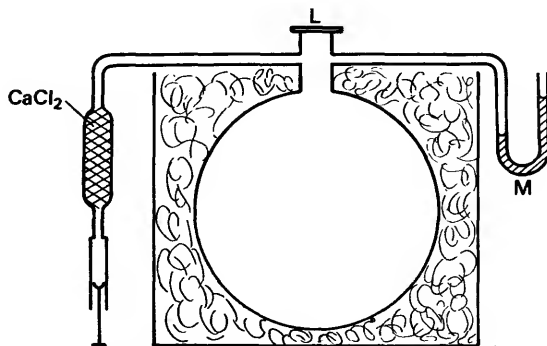


FIG. 10.22. Clément and Désormes experiment.

carboy is well lagged to minimize the exchange of heat with its surroundings. It is attached to a manometer M, and, via a drying-tube, to a bicycle pump. Its mouth has a large and well-fitting, flap-like, lid, L. Air is blown in until its pressure is a little above atmospheric, and time is allowed for the gas to settle down to room temperature. When it has done so the manometer reading becomes steady, and the pressure p_1 of the gas is recorded. The flap-valve is now sharply opened and closed. The gas makes an adiabatic expansion, and its pressure p_2 is immediately read. With the flap still closed, the gas is then left; it

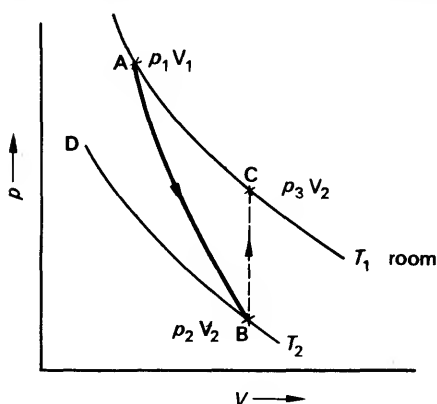


FIG. 10.23. Theory of Clément and Désormes.

gradually returns to room temperature, absolute temperature T_1 , at constant volume, and its pressure rises to p_3 .

These changes are shown in Fig. 10.23. Since some gas escapes in the expansion, we must consider unit mass. Its state at the start of the experiment is represented by the point A on the isothermal for T_1 , its volume being V_1 . B represents the end of the adiabatic, when the gas has cooled to T_2 , and expanded to V_2 per unit mass. DB is the isothermal for T_2 . BC represents the return to room temperature. For the adiabatic AB, we have

$$p_1 V_1^\gamma = p_2 V_2^\gamma,$$

$$\text{or} \quad \frac{p_1}{p_2} = \left(\frac{V_2}{V_1} \right)^\gamma \quad (33)$$

After the gas has returned to room temperature, its representative point C lies on the same isothermal as A; therefore

$$p_3 V_2 = p_1 V_1,$$

$$\text{or} \quad \frac{p_1}{p_3} = \frac{V_2}{V_1}.$$

From equation (33), therefore,

$$\frac{p_1}{p_2} = \left(\frac{p_1}{p_3} \right)^\gamma$$

whence $\log_e p_1 - \log_e p_2 = \gamma(\log_e p_1 - \log_e p_3)$

and

$$\gamma = \frac{\log_e p_1 - \log_e p_2}{\log_e p_1 - \log_e p_3}$$

If h_1 is the difference in levels of M corresponding to the pressure

p_1 , and h_2 is the final difference in levels, corresponding to the pressure p_3 , it can be shown that, to a good approximation, the formula for γ reduces to

$$\gamma = \frac{h_1}{h_1 - h_2}.$$

A light oil is used in the manometer M.

VACUUM PUMPS AND GAUGES

The Filter Pump

The simplest pump for evacuating a vessel is the filter pump, so-called because it is used for speeding-up filtration (Fig. 10.24). It consists of a nozzle N surrounded by a chamber C; water rushes from the nozzle and out of the chamber at the bottom. The layer of air around the jet is dragged along with it, and carried out of the chamber. The lowest pressure which this pump can produce is the saturation vapour-pressure of water, about 15 mm mercury at 18°C. (Chapter 14). It cannot produce what we would nowadays call a 'good vacuum'.

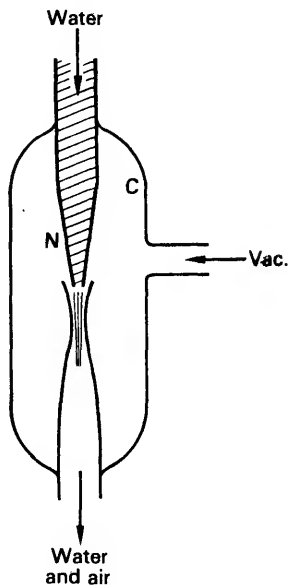


FIG. 10.24. Filter pump.

The Piston Pump

A piston-type air pump is similar to the common water pump, but more accurately made. Its plunger has a greased leather washer, W in Fig. 10.25, and its valves F_1 and F_2 are flaps of oiled silk.

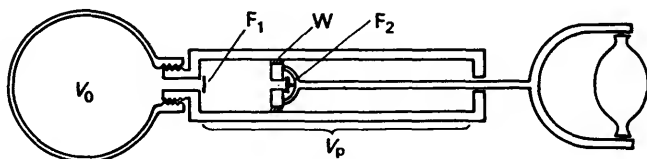


FIG. 10.25. Piston-type vacuum pump.

To develop a simple theory of the pump, we first assume that, when the piston is pushed right in, there is no space between it and the bottom of the barrel. We suppose that the pump is connected to a vessel of volume V_0 , that the piston displaces a volume V_p , and that the pressure in the vessel is p_0 when the piston is right in, at the start of the evacuation. When the piston is pulled right out, the volume of the air originally in the vessel increases to $V_0 + V_p$. Since the action

is slow, we may assume that the expansion is isothermal; the pressure p_1 after expansion is therefore given by

$$p_1(V_0 + V_P) = p_0 V_0,$$

whence

$$p_1 = \left(\frac{V_0}{V_0 + V_P} \right) p_0.$$

When the piston is pushed in again, the valve F_1 closes, and the air in the vessel remains at the pressure p_1 . The second out-stroke then reduces the pressure to

$$p_2 = \left(\frac{V_0}{V_0 + V_P} \right) p_1 = \left(\frac{V_0}{V_0 + V_P} \right)^2 p_0.$$

Similarly, after n strokes, the pressure is reduced to

$$p_n = \left(\frac{V_0}{V_0 + V_P} \right)^n p_0.$$

According to this theory, the final pressure tends to zero as the number of strokes tends to infinity. In practice, however, a pump has a limiting pressure. This is due to the fact that the piston can never in practice be brought right down to the valve F_1 , so that there is a residual volume, or dead-space, v , between the piston and the bottom of the barrel. Air can escape through the valve F_2 only when the pressure in the volume v is greater than atmospheric. And air can pass from the vessel through F_1 only when the pressure in the barrel is less than the pressure in the vessel. Thus the limiting pressure, p_∞ , is the pressure which $v \text{ cm}^3$ of gas, at atmospheric pressure, exert when expanded to $V_P \text{ cm}^3$. That is to say

$$p_\infty V_P = p_{\text{atmos.}} v,$$

whence

$$p_\infty = \frac{v}{V_P} p_{\text{atmos.}}$$

The ratio v/V_P may be about $1/1000$, so that p_∞ is about 1 mm mercury. Piston pumps of more elaborate designs can give better vacua than this, about 0.01 mm mercury; they have all been made obsolete, however, by the rotary pump.

Rotary Pumps

Fig. 10.26 illustrates one form of rotary vacuum pump. It consists of a rotor, R, which turns eccentrically in a casing, C, being a close fit at the ends and along the line A. The rotor carries two scraping blades B_1 , B_2 , separated by a strong spring. The vessel to be evacuated is connected to the inlet port I, and the outlet port O is fitted with a valve N. As the rotor turns, the volume V_1 increases, so that air expands from the vessel into the pump. When the blades are in the positions shown, the air in the space V_2 is being compressed; when its pressure rises to atmospheric, the valve N opens, and the air passes out through O. As the blade B_2 crosses the seat of the valve N, the valve closes

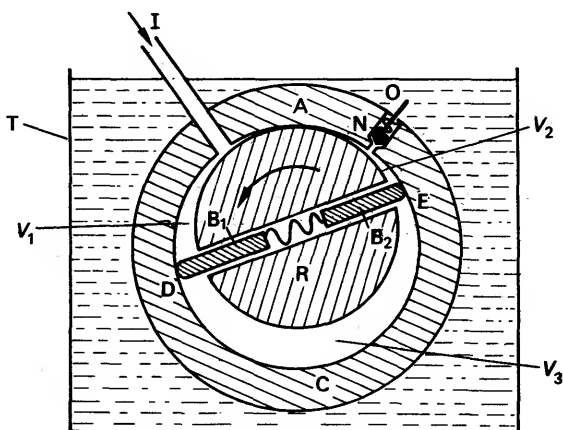


FIG. 10.26. Rotary vacuum pump.

because the air in V_3 is below atmospheric pressure. Thus atmospheric air cannot blow back into the pump. The lines of contact, A, D, E are made airtight by a film of oil. All the working parts of the pump are enclosed in a tank T containing oil. When the pump is at rest, the oil seeps back through the outlet valve and fills the working space; but the first revolutions of the pump sweep out the excess oil, and leave just the necessary film over the metal surfaces.

A single rotary pump will give an ultimate pressure of about 0.01 mm mercury. Very often two pumps are housed in the same tank of oil, and driven off the same shaft; they are connected in cascade, and may give an ultimate pressure of less than 0.001 mm mercury.

The MacLeod Gauge

The MacLeod Gauge is a gauge used for measuring pressures below a few mm mercury; these are pressures which cannot be measured accurately on a U-tube manometer. It consists of a bulb B, connected to a mercury reservoir M and terminated in a capillary tube T (Fig. 10.27 (a)); just below the bulb a branch-tube, P, leads to the vacuum system under test. It also carries a branch-tube, D, which is a capillary of the same bore as T. A millimetre scale, S, lies underneath T and D.

To measure the pressure in a vacuum system, the reservoir M is lowered until the mercury falls below the branch-point C. The air in B is then at the unknown pressure, p . The reservoir is now slowly raised. As soon as the mercury closes the branch at C, the air in B starts to be compressed. M is raised until the mercury in B just reaches the foot of the capillary T. The height h of the mercury in D, above that in B, is then measured. The purpose of having equal bores for T and D is to equalize surface tension effects in each.

If the pressure p is expressed in mm mercury then the pressure of the air trapped in T is $p + h$. The volume of this air is the volume v of the capillary T. At the moment when the mercury passed the point C, this air had the pressure p , and the volume $V + v$, where V is the

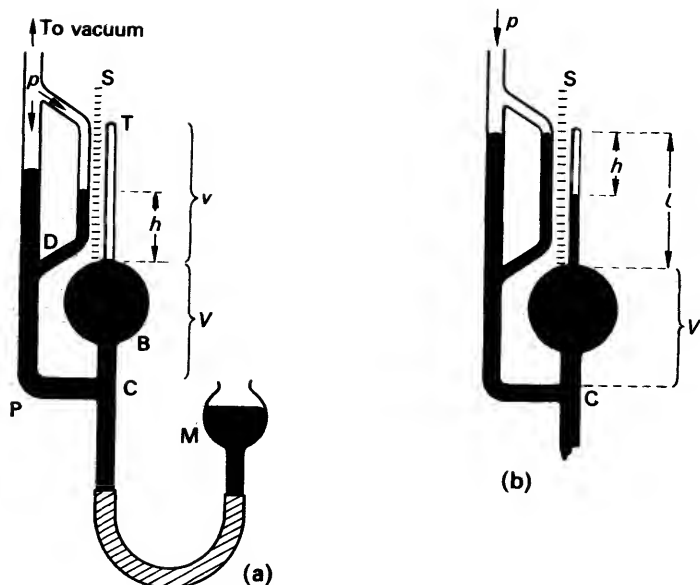


FIG. 10.27. MacLeod gauge.

volume between C and the base of T. The compression is slow enough to be isothermal, so that

$$p(V+v) = (p+h)v$$

Hence

$$pV = hv,$$

and

$$p = \frac{v}{V}h.$$

Another way of using the gauge is to raise the reservoir M until the mercury in D is level with the top of T (Fig. 10.27 (b)). Then if s is the cross-section of T, the volume of trapped air is hs . And if l is the whole length of T, its volume is ls . Therefore, as before,

$$p(V+ls) = (p+h)hs,$$

or

$$p(V+ls-hs) = h^2s,$$

whence

$$p = \frac{h^2s}{V+(l-h)s}.$$

The term $(l-h)s$ in the denominator is usually negligible compared with V , so that

$$p \simeq \frac{h^2s}{V}.$$

Because p is proportional to h^2 , the gauge can cover a wider range of pressures when it is used in this way, than when it is used in the way first described. But for the same reason it is less accurate. In practice

the second way is generally chosen, and the scale S is calibrated to read the pressure p directly.

EXAMPLES

1. The density of a gas is 10775 kg m^{-3} at 27°C and 10^5 N m^{-2} pressure and its specific heat capacity at constant pressure is $0.846 \text{ kJ kg}^{-1} \text{ K}^{-1}$. Find the ratio of its specific heat capacity at constant pressure to that at constant volume.

The gas constant per kg of gas is given by

$$R = \frac{pV}{T} = \frac{10^5 \times 1}{1.775 \times 300} \text{ J kg}^{-1} \text{ K}^{-1}$$

since $V = 1 \text{ m}^3/1.775$, $T = 273 + 27 = 300 \text{ K}$. Converting J to kJ,

$$\therefore R = \frac{10^5 \times 1}{1.775 \times 300 \times 1000} \text{ kJ kg}^{-1} \text{ K}^{-1}.$$

Now

$$c_p - c_v = R$$

$$\therefore 0.846 - c_v = \frac{10^5 \times 1}{1.775 \times 300 \times 1000} = 0.188$$

$$\therefore c_v = 0.846 - 0.188 = 0.658 \text{ kJ kg}^{-1} \text{ K}^{-1}$$

$$\therefore \gamma = \frac{c_p}{c_v} = \frac{0.846}{0.658} = 1.29.$$

This value for γ suggests that the gas is polyatomic (see p. 246).

2. An ideal gas at 17°C has a pressure of 760 mm mercury, and is compressed (i) isothermally, (ii) adiabatically until its volume is halved, in each case reversibly. Calculate in each case the final pressure and temperature of the gas, assuming $c_p = 2.1$, $c_v = 1.5 \text{ kJ kg}^{-1} \text{ K}^{-1}$.

(i) Isothermally, $pV = \text{constant}$.

$$\therefore p \times \frac{V}{2} = 760 \times V$$

$$\therefore p = 1520 \text{ mm mercury.}$$

The temperature is constant at 17°C .

(ii) Adiabatically, $pV^\gamma = \text{constant}$, and $\gamma = 2.1/1.5 = 1.4$.

$$\therefore p \times \left(\frac{V}{2}\right)^{1.4} = 760 \times V^{1.4}$$

$$\therefore p = 760 \times 2^{1.4} = 2010 \text{ mm mercury.}$$

Since $TV^{\gamma-1} = \text{constant}$,

$$\therefore T \times \left(\frac{V}{2}\right)^{0.4} = (273 + 17) \times V^{0.4}$$

$$\therefore T = 290 \times 2^{0.4} = 383 \text{ K.}$$

$$\therefore \text{temperature} = 110^\circ\text{C.}$$

3. State the laws of gases usually associated with the names of Boyle, Charles, Dalton and Graham. Two gas containers with volumes of 100 cm^3 and 1000 cm^3 respectively are connected by a tube of negligible volume, and contain air at a

pressure of 1000 mm mercury. If the temperature of both vessels is originally 0°C , how much air will pass through the connecting tube when the temperature of the smaller is raised to 100°C ? Give your answer in cm^3 measured at 0°C and 760 mm mercury. (L.)

First part. Boyle, Charles, Dalton, Graham, see text.

Second part. The pressure is 1000 mm mercury when the temperature is 0°C (273 K). Let the density of air under these conditions be ρ_1 . Let the volumes of the large and small vessels be V and V^1 ; then the mass of air in the two vessels is

$$M = (V + V^1)\rho_1 = (1000 + 100)\rho_1 = 1100\rho_1 \quad \text{. (i)}$$

When the smaller vessel is heated, the pressure throughout the system rises to p , say. Let ρ_2 be the density of the air in the smaller vessel; then, by equation (9), p. 227:

$$\frac{p}{\rho_2} = R \times 373; \quad \frac{1000}{\rho_1} = R \times 273 \quad \text{. (ii)}$$

$$\therefore \frac{\rho_2}{\rho_1} = \frac{273}{373} \times \frac{p}{1000}$$

$$\therefore \rho_2 = \frac{273}{373} \times \frac{p}{1000} \rho_1$$

In the larger vessel, the temperature of the air does not change; therefore the density of the air in the larger vessel, ρ_3 , is

$$\rho_3 = \frac{p}{1000} \rho_1$$

The total mass of air, which is unchanged, is therefore

$$\begin{aligned} M &= V\rho_3 + V^1\rho_2 = 1000\rho_3 + 100\rho_2 \\ &= 1000 \frac{p\rho_1}{1000} + 100 \frac{273}{373} \frac{p\rho_1}{1000} \\ &= 1 + \left(\frac{273}{3730} \right) p\rho_1 \end{aligned}$$

Hence, by equation (i),

$$1100\rho_1 = \left(1 + \frac{273}{3730} \right) p\rho_1,$$

and

$$p = \frac{3730 \times 1100}{4003} = 1025 \text{ mm mercury.}$$

The mass which flows out of the smaller vessel is

$$\begin{aligned} m &= V^1(\rho_1 - \rho_2) = 100\rho_1 \left(1 - \frac{\rho_2}{\rho_1} \right) \\ &= 100\rho_1 \left(1 - \frac{273}{373} \times \frac{p}{1000} \right) \\ m &= 100\rho_1 \left(1 - \frac{273}{373} \times \frac{1025}{1000} \right) \quad \text{. (iii)} \end{aligned}$$

The volume of this mass, at 0°C and 760 mm mercury, is

$$V = \frac{m}{\rho_4},$$

where ρ_4 is the density of air at this temperature and pressure.

From the equation of state,

$$\frac{760}{\rho_4} = R \times 273;$$

therefore
$$\frac{\rho_1}{\rho_4} = \frac{1000}{760},$$

or
$$\frac{1}{\rho_4} = \frac{100}{76\rho_1}.$$

Hence, by (iii),
$$V = \frac{m}{\rho_4} = \frac{100 \times 100}{76} \left(1 - \frac{273}{373} \times \frac{1025}{1000} \right)$$

$$= 33 \text{ cm}^3.$$

4. Distinguish between *isothermal* and *adiabatic* changes. Show that for an ideal gas the curves relating pressure and volume for an adiabatic change have a greater slope than those for an isothermal change, at the same pressure.

A quantity of oxygen is compressed isothermally until its pressure is doubled. It is then allowed to expand adiabatically until its original volume is restored. Find the final pressure in terms of the initial pressure. (The ratio of the specific heat capacities of oxygen is to be taken at 1.40.) (L.)

First part. An isothermal change is one made at constant temperature; an adiabatic change is one made at constant heat, that is, no heat enters or leaves the system concerned.

For a reversible isothermal change, $pV = \kappa$, or $p = \kappa/V$. By differentiation, the slope, dp/dV , $= -\kappa/V^2 = -pV/V^2 = -p/V$.

For a reversible adiabatic change, $pV^\gamma = c$, or $p = c/V^\gamma$. By differentiation, we find the slope, dp/dV , $= -\gamma p/V$.

\therefore ratio of adiabatic slope to isothermal slope $= \gamma$.

Since γ is always greater than 1, the adiabatic slope is greater than the isothermal slope.

Second part. Let p_0 , V_0 = the original pressure and volume of the oxygen.

Since $pV = \text{constant}$ for an isothermal change,

$$\therefore \text{new volume} = \frac{V_0}{2} \text{ when new pressure is } 2p_0.$$

Suppose the gas expands adiabatically to its volume V_0 , when the pressure is p .

Then
$$p \times V_0^{1.4} = 2p_0 \times \left(\frac{V_0}{2} \right)^{1.4}$$

$$\therefore p = 2p_0 \times \left(\frac{1}{2} \right)^{1.4} = 0.8 p_0.$$

5. Derive an expression for the difference between the specific heat capacities of an ideal gas and discuss the significance of the ratio of these two specific heat capacities for real gases.

Assuming that the ratio of the specific heat capacities of hydrogen is 1.41 and that its density at s.t.p. is 0.0900 kg m^{-3} , find a value for its specific heat capacity at constant volume in $\text{J kg}^{-1} \text{ K}^{-1}$.

What explanation can you suggest for the small difference between the specific heat capacities of a solid? (Standard atmospheric pressure $= 1.013 \times 10^5 \text{ N m}^{-2}$.) (N.)

First part. The expression required is $c_p - c_v = R$, discussed on p. 245. The

ratio, γ , of these two specific heats $= 1 + \frac{2}{n}$ on the kinetic theory of gases, where n is the number of degrees of freedom of the molecules. For a monatomic gas $n = 3$, so that $\gamma = 1.66$; for a diatomic gas $n = 5$ usually, so that $\gamma = 1.4$; for triatomic gases γ is less than 1.4, e.g. 1.29. Thus γ gives information about the number of atoms in the molecule of the gas.

Second part. $c_p - c_v = R$, where R may be in $\text{J kg}^{-1} \text{K}^{-1}$ and c_p, c_v are in the same units. Since 0.09 kg occupies 1 m^3 and $p = 1.013 \times 10^5 \text{ newton m}^{-2}$, then, from $pV = RT$,

$$R = \frac{pV}{T} = \frac{(1.013 \times 10^5) \times 1}{273 \times 0.09} \text{ J kg}^{-1} \text{K}^{-1} \\ = 4.12 \text{ kJ kg}^{-1} \text{K}^{-1}$$

$$\therefore c_p - c_v = 4.12 \quad \quad \quad (i)$$

$$\text{But} \quad \frac{c_p}{c_v} = 1.41 \quad \quad \quad (ii)$$

$$\therefore c_p = 1.41 c_v. \text{ Substituting for } c_p \text{ in (i),}$$

$$\therefore 1.41 c_v - c_v = 4.12 = 0.41 c_v.$$

$$\therefore c_v = \frac{4.12}{0.41} = 10.0 \text{ kJ kg}^{-1} \text{K}^{-1}.$$

Third part. The difference in the specific heat capacities of a solid is proportional to the external work done in expansion. But the expansion of a solid is small. Consequently the difference in specific heat capacities of the solid is small.

EXERCISES 10

Gas Laws—Specific Heat Capacities of Gases

1. State *Boyle's law* and *Charles' law*, and show how they lead to the gas equation $PV = RT$. Describe an experiment you would perform to measure the thermal expansion coefficient of dry air.

What volume of liquid oxygen (density 1140 kg m^{-3}) may be made by liquefying completely the contents of a cylinder of gaseous oxygen containing 100 litres of oxygen at 120 atmospheres pressure and 20°C ? Assume that oxygen behaves as an ideal gas in this latter region of pressure and temperature.

[1 atmosphere $= 1.01 \times 10^5 \text{ newton metre}^{-2}$; gas constant $= 8.31 \text{ joule mol}^{-1} \text{K}^{-1}$; molecular weight of oxygen $= 32.0$.] (O. & C.)

2. Give brief accounts of experiments which illustrate the relationship between the volume of a fixed mass of gas and (a) the pressure it exerts at a fixed temperature, (b) the temperature on a centigrade mercury thermometer at a fixed pressure. State the two 'laws' which summarise the results.

A gas cylinder contains 6400 g of oxygen at a pressure of 5 atmospheres. An exactly similar cylinder contains 4200 g of nitrogen at the same temperature. What is the pressure on the nitrogen? (Molecular weights: oxygen $= 32$, nitrogen $= 28$; assume that each behaves as a perfect gas.) (O. & C.)

3. State Boyle's Law.

Describe how you would verify the law for dry air over a range of pressures from 0.5 to 1.5 atmospheres. Would the form of the apparatus you describe be suitable if the working range of pressure was 0.5 to 10 atmospheres? Give reasons for your answer.

Two glass vessels of equal volume are joined by a tube, the volume of which may be neglected. The whole is sealed and contains air at S.T.P. If one vessel is

placed in boiling water at 100°C and the other is placed in melting ice, what will be the resultant pressure of the air? (N.)

4. The formula $pv = mrT$ is often used to describe the relationship between the pressure p , volume v , and temperature T of a mass m of a gas, r being a constant. Referring in particular to the experimental evidence how do you justify (a) the use of this formula, (b) the usual method of calculating T from the temperature t of the gas on the centigrade (Celsius) scale?

Two vessels each of capacity 1.00 litre are connected by a tube of negligible volume. Together they contain 0.342 g of helium at a pressure of 80 cm of mercury and temperature 27°C . Calculate (i) a value for the constant r for helium, (ii) the pressure developed in the apparatus if one vessel is cooled to 0°C and the other heated to 100°C , assuming that the capacity of each vessel is unchanged. (N.)

5. State Boyle's law and Charles' law and show how they may be combined to give the equation of state of an ideal gas.

Two glass bulbs of equal volume are joined by a narrow tube and are filled with a gas at s.t.p. When one bulb is kept in melting ice and the other is placed in a hot bath, the new pressure is 87.76 cm mercury. Calculate the temperature of the bath. (L.)

6. Describe experiments in which the relation between the pressure and volume of a gas has been investigated at constant temperature over a wide range of pressure. Sketch the form of the isothermal curves obtained.

Explain briefly how far van der Waals' equation accounts for the form of these isothermals. (L.)

7. Describe, with a diagram, the essential features of an experiment to study the departure of a real gas from ideal gas behaviour.

Give freehand, labelled sketches of the graphs you would expect to obtain on plotting (a) pressure P against volume V , (b) PV against P for such a gas at its critical temperature and at one temperature above and one below the critical temperature.

Explain van der Waals' attempt to produce an equation of state which would describe the behaviour of real gases.

Show that van der Waals' equation is consistent with the statement that all gases approach ideal gas behaviour at low pressures. (O. & C.)

8. State Boyle's Law and describe how you would attempt to discover whether air shows any deviations from the law. Draw an approximate set of curves to show the way in which a gas deviates from Boyle's Law in the region close to where the gas liquefies.

Two one-litre flasks are joined by a closed tap and the whole is held at a constant temperature of 50°C . One flask is evacuated and the other contains air, water vapour, and a small quantity of liquid water. The total pressure in the latter flask is 200 mm Hg. The tap is then opened, and the system is allowed to reach equilibrium, when some liquid water remains. Assuming that air obeys Boyle's law, find the final pressure in the flasks.

[Vapour pressure of water at $50^{\circ}\text{C} = 93 \text{ mmHg.}]$ (O. & C.)

9. Explain why the specific heat capacity of a gas is greater if it is allowed to expand while being heated than if the volume is kept constant. Discuss whether it is possible for the specific heat capacity of a gas to be zero.

When 1 g of water at 100°C is converted into steam at the same temperature 2264 J must be supplied. How much of this energy is used in forcing back the atmosphere? Explain what happens to the remainder of the energy. [1 g of water 100°C occupies 1 cm^3 . 1 g of steam at 100°C and 76 cm of mercury occupies 1601 cm^3 . Density of mercury = 13600 kg m^{-3} .] (C.)

10. Define *heat capacity* and *specific heat capacity*.

Describe an experiment to determine the specific heat capacity of a gas *either* at constant volume *or* at constant pressure. Point out likely sources of error and indicate how they may be minimized.

Explain why it is necessary to specify the condition of constant pressure or constant volume. (L.)

11. Explain why, when quoting the specific heat of a gas, it is necessary to specify the conditions under which the change of temperature occurs. What conditions are normally specified?

A vessel of capacity 10 litres contains 130 g of a gas at 20°C and 10 atmospheres pressure. 8000 joule of heat energy are suddenly released in the gas and raise the pressure to 14 atmospheres. Assuming no loss of heat to the vessel, and ideal gas behaviour, calculate the specific heat of the gas under these conditions.

In a second experiment the same mass of gas, under the same initial conditions; is heated through the same rise in temperature while it is allowed to expand slowly so that the pressure remains constant. What fraction of the heat energy supplied in this case is used in doing external work? Take 1 atmosphere = 10^5 newton metre $^{-2}$. (O. & C.)

12. The two specific heat capacities in $\text{kJ kg}^{-1} \text{K}^{-1}$ units for argon are 0.521 and 0.313 and for air are 1.012 and 0.722. Explain these statements and discuss their significance in relation to (a) the atomicity of the molecules of the two gases, (b) the relative values of the adiabatic elasticities of argon and air.

Describe an experiment to verify *one* of the above values of specific heat capacities. (L.)

13. A litre of air, initially at 20°C and at 76.0 cm of mercury pressure, is heated at constant pressure until its volume is doubled. Find (a) the final temperature, (b) the external work done by the air in expanding, (c) the quantity of heat supplied.

[Assume that the density of air at s.t.p. is 1.293 kg m^{-3} and that the specific heat capacity of air at constant volume is $0.714 \text{ kJ kg}^{-1} \text{K}^{-1}$.] (L.)

14. Describe an experiment to determine the specific heat capacity of water, at about 15°C , deriving from first principles any equations used.

Deduce an expression for the difference between the specific heat capacities of an ideal gas. If the specific heat capacity of air at constant pressure is $1.013 \text{ kJ kg}^{-1} \text{K}^{-1}$ and the density at s.t.p. is 1.29 kg m^{-3} estimate a value for the specific heat capacity of air at constant volume. [Assume the density of mercury at 0°C to be 13600 kg m^{-3} .] (L.)

15. Distinguish between an *isothermal change* and an *adiabatic change*. In each instance state, for a reversible change of an ideal gas, the relation between pressure and volume.

A mass of air occupying initially a volume 2000 cm^3 at a pressure of 76.0 cm of mercury and a temperature of 20.0°C is expanded adiabatically and reversibly to twice its volume, and then compressed isothermally and reversibly to a volume of 3000 cm^3 . Find the final temperature and pressure, assuming the ratio of the specific heat capacities of air to be 1.40. (L.)

16. Explain why the specific heat of a gas at constant pressure is different from that at constant volume.

The density of an ideal gas is 1.60 kg m^{-3} at 27°C and 1.00×10^5 newton metre $^{-2}$ pressure and its specific heat capacity at constant volume is $0.312 \text{ kJ kg}^{-1} \text{K}^{-1}$. Find the ratio of the specific heat capacity at constant pressure to that at constant volume. Point out any significance to be attached to the result. (N.)

17. Explain the meaning of the terms *isothermal*, *adiabatic*. What is the importance of the ratio of the specific heat capacity of an ideal gas?

Air initially at 27°C and at 75 cm of mercury pressure is compressed isothermally until its volume is halved. It is then expanded adiabatically until its original volume is recovered. Assuming the changes to be reversible, find the final pressure and temperature.

[Take the ratio of the specific heat capacities of air as 1.40.] (L.)

Kinetic Theory of Gases

18. Explain what is meant by the *root mean square velocity* of the molecules of a gas. Use the concepts of the elementary kinetic theory of gases to derive an expression for the root mean square velocity of the molecules in terms of the pressure and density of the gas.

Assuming the density of nitrogen at s.t.p. to be 1.251 kg m^{-3} , find the root mean square velocity of nitrogen molecules at 127°C . (L.)

19. State the postulates on which the simple kinetic theory of gases is based. What modifications are made to the postulates in dealing with real gases? How are these modifications represented in van der Waals' equation? (N.)

20. State the assumptions that are made in the kinetic theory of gases and derive an expression for the pressure exerted by a gas which conforms to these assumptions, in terms of its density (ρ) and the mean square velocity (c^2) of its molecules.

Show (a) how *temperature* may be interpreted in terms of the theory, (b) how the theory accounts for Dalton's law of partial pressures. (L.)

21. Calculate the pressure in mm of mercury exerted by hydrogen gas if the number of molecules per cm^3 is 6.80×10^{15} and the root mean square speed of the molecules is $1.90 \times 10^3\text{ m s}^{-1}$. Comment on the effect of a pressure of this magnitude (a) above the mercury in a barometer tube; (b) in a cathode ray tube. (Avogadro's Number = 6.02×10^{23} . Molecular weight of hydrogen = 2.02.) (N.)

22. Use a simple treatment of the kinetic theory of gases, stating any assumptions you make, to derive an expression for the pressure exerted by a gas on the walls of its container. Thence deduce a value for the root mean square speed of thermal agitation of the molecules of helium in a vessel at 0°C . (Density of helium at s.t.p. = 0.1785 kg m^{-3} ; 1 atmosphere = $1.013 \times 10^5\text{ N m}^{-2}$.)

If the total translational kinetic energy of all the molecules of helium in the vessel is $5 \times 10^{-6}\text{ joule}$, what is the temperature in another vessel which contains twice the mass of helium and in which the total kinetic energy is 10^{-5} joule ? (Assume that helium behaves as a perfect gas.) (O. & C.)

23. Explain the meaning of the terms *ideal gas* and *molecule*.

What properties of a gas such as carbon dioxide distinguish it from an ideal gas and how may these differences from 'ideal' be demonstrated experimentally?

According to simple kinetic theory the pressure exerted by a gas of density ρ is $\frac{1}{3}\rho c^2$ where c^2 is the mean square molecular velocity. Show how this relation may be correlated with the equation of state for an ideal gas, $PV = RT$, explaining clearly what further assumptions you have to make. (O. & C.)

24. Derive an expression for the pressure (p) exerted by a mass of gas in terms of its molecular velocities, stating the assumptions made. What further assumption regarding the absolute temperature (T) of the gas is necessary to show that the expression is consistent with the equation $pv = kT$ where v is the volume of the mass of gas and k a constant? Which of the assumptions referred to did van der Waals modify to bring the equation $pv = kT$ more closely in agreement with the behaviour of real gases and what equation did he deduce? (L.)

25. Explain the following in terms of the simple kinetic theory without mathematical treatment:

(a) A gas fills any container in which it is placed and exerts a pressure on the walls of the container.

(b) The pressure of a gas rises if its temperature is increased without the mass and volume being changed.

(c) The temperature of a gas rises if it is compressed in a vessel from which heat cannot escape.

(d) The pressure in an oxygen cylinder falls continuously as the gas is taken from it, while the pressure in a cylinder containing chlorine remains constant until very nearly all the chlorine has been used. The contents of the cylinder are kept at room temperature in both cases.

[Critical temperature of chlorine = 146°C .] (C.)

26. (a) State the assumptions of the kinetic theory of gases. How does the theory represent the temperature of a gas, and how does it account for the fact that a gas exerts a pressure on the walls of its container? (The derivation of an expression for the pressure is NOT required.)

(b) Near 0 K the specific heat capacity of silver, c , is not constant, but obeys the relation

$$c = \alpha T^3 + \beta T,$$

where T is the absolute temperature and α and β are constants typical of silver given by $\alpha = 15.12 \times 10^{-7} \text{ kJ kg}^{-1} \text{ deg}^{-4}$ and $\beta = 5.88 \times 10^{-6} \text{ kJ kg}^{-1} \text{ deg}^{-1}$. By means of a graph, or otherwise, find the heat required to raise the temperature of 5.0 g of silver from 1 K to 20 K. (C.)

27. Explain in terms of the kinetic theory what happens to the energy supplied to a gas when it is heated (a) at constant volume, (b) at constant pressure.

Deduce the total kinetic energy of the molecules in 1 g of an ideal gas at 0°C if its specific heat capacity at constant volume is $0.60 \text{ kJ kg}^{-1} \text{ K}^{-1}$.

An iron rod 1 metre long is heated without being allowed to expand lengthwise. When the temperature has been raised by 500°C the rod exerts a force of 1.2×10^4 newtons on the walls preventing its expansion. How much work could be obtained if it were possible to maintain it at this temperature and allow it to expand gradually until free from stress? [Linear expansivity of iron = $1.0 \times 10^{-5} \text{ K}^{-1}$.] (C.)

chapter eleven

Thermal expansion

IN this chapter we shall discuss the thermal expansion of solids and liquids.

SOLIDS

Linear Expansion

Most solids increase in length when they are warmed. Fig. 11.1 shows a simple apparatus with which we can measure the linear expansion of a metal tube A. We first measure the length of the tube, l_1 , at room temperature, θ_1 ; then we screw the spherometer S against the end of the tube and take its reading, S_1 . We next heat the tube by passing steam through it. At intervals we re-adjust the spherometer; when its reading becomes constant, the temperature of the rod is steady. We measure the temperature θ_2 on the thermometer B, and take the new reading of the spherometer, S_2 . The expansion of the tube is

$$e = S_2 - S_1.$$

The increase in length, λ , of unit length of the material for one degree temperature rise is then given by

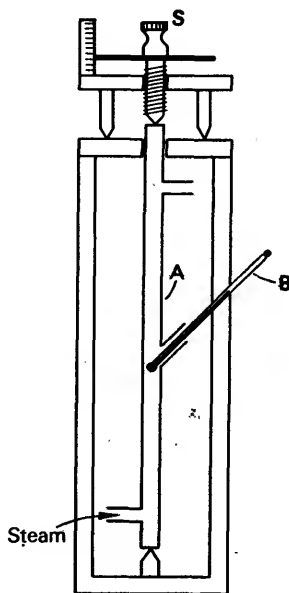


FIG. 11.1. Linear expansivity..

$$\lambda = \frac{\text{expansion}}{\text{original length} \times \text{temperature rise}} = \frac{e}{l_1(\theta_2 - \theta_1)}$$

The quantity λ is called the *mean linear expansivity* of the metal, over the range θ_1 to θ_2 . If this range is not too great—say less than 100°C —the quantity λ may, to a first approximation, be taken as constant.

The linear expansivity of a solid, like the pressure and volume coefficients of a gas, has the unit deg C^{-1} or $^\circ\text{K}^{-1}$ in SI units; its dimensions are

$$[\lambda] = \frac{[\text{length}]}{[\text{length}] \times [\text{temp.}]} = [\text{temp.}]^{-1}.$$

From the definition of λ , we can estimate the new length of a rod, l_2 , at a temperature θ_2 from the equation

$$l_2 = l_1 \{1 + \lambda(\theta_2 - \theta_1)\}, \quad (1)$$

where l_1 is the length of the rod at the temperature θ_1 , and λ is the mean value over a range which includes θ_2 and θ_1 .

For accurate work, however, the length of a solid at a temperature θ must be represented by an equation of the form

$$l = l_0(1 + a\theta + b\theta^2 + c\theta^3 + \dots), \quad (2)$$

where l_0 is the length at 0°C , and a, b, c are constants. The constant a is of the same order of magnitude as the mean coefficient λ ; the other constants are smaller.

MEAN LINEAR EXPANSIVITY

(Near room temperature)

Substance	λ, K^{-1} $\times 10^{-6}$	Substance	λ, K^{-1} $\times 10^{-6}$
Copper	17	Bakelite	22
Iron	12	Brick	9.5
Brass	19	Glass (soda)	8.5
Nickel	13	Quartz (fused) ($0-30^\circ\text{C}$) .	0.42
Platinum	9	Pine—across grain	c. 0.34
Invar (36% nickel-steel) .	c. 0.1	Pyrex	3

When a solid is subjected to small changes of temperature, about a mean value θ , its linear expansivity λ_θ in the neighbourhood of θ may be defined by the equation

$$\lambda_\theta = \frac{1}{l} \frac{dl}{d\theta},$$

where l is the length of the bar at the temperature θ . The following table shows how the coefficient varies with temperature.

VALUES OF λ_θ COPPER

θ	-87	0	100	400	600	$^\circ\text{C}$
λ_θ	14.1	16.1	16.9	19.3	20.9	$\times 10^{-6} \text{K}^{-1}$

Accurate Measurement of Expansion

An instrument for accurately measuring the length of a bar, at a controlled temperature, is called a *comparator* (Fig. 11.2). It consists of two microscopes, M_1, M_2 , rigidly attached to steel or concrete pillars P_1, P_2 . Between the pillars are rails R_1, R_2 , carrying water-baths such as B. One of these baths contains the bar under test, X, which has scratches near its ends; the scratches are nominally a metre apart. Another water bath contains a substandard metre. The eyepieces of the microscopes are fitted with cross-webs carried on micrometer screws, m_1, m_2 .

First the substandard metre is run under the microscopes, and the temperature of its bath is adjusted to that at which the bar was calibrated (usually about 18°C).

When the temperature of the bar is steady, the eyepiece webs are adjusted to intersect the scratches on its ends (Fig. 11.2 (b)), and their micrometers are read. The distance between the cross-webs is then 1 metre.

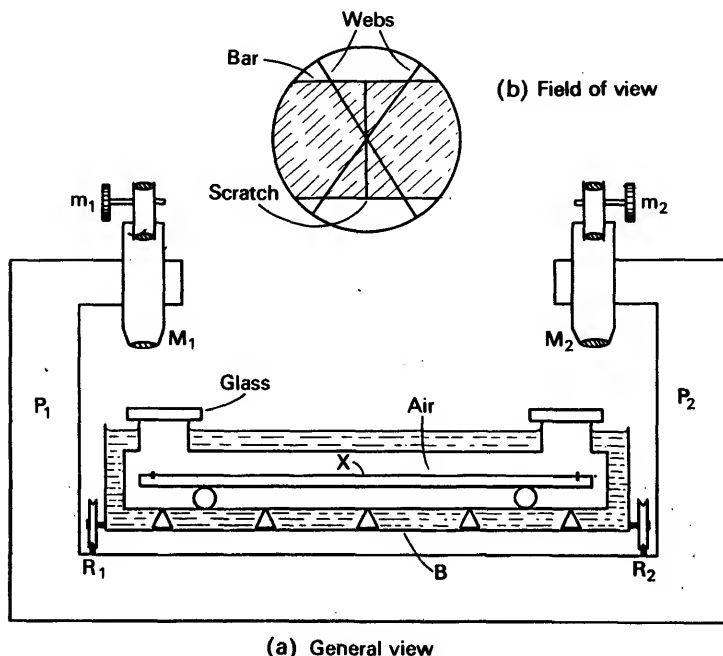


FIG. 11.2. The comparator.

The substandard is now removed, and the unknown bar put in its place; the temperature of the bar is brought to 0°C by filling its bath with ice-water. When the temperature of the bar is steady, the eyepiece webs are re-adjusted to intersect the scratches on its ends, and their micrometers are read. If the right-hand web has been shifted x mm to the right, and the left-hand y mm, also to the right, then the length of the bar at 0°C is

$$l_0 = 1 \text{ metre} + (x - y) \text{ mm.}$$

The bath is now warmed to say 10°C , and the length of the bar again measured. In this way the length can be measured at small intervals of temperature, and the mean linear expansivity, or the coefficients a , b , c in equation (2), can be determined.

Expansion at High Temperatures

Fig. 11.3 illustrates the principle of a method for measuring the expansion of a solid at high temperatures. A is a tube of fused silica,

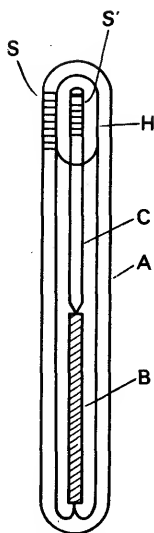


FIG. 11.3. Apparatus for measuring expansion at high temperatures.

having a scale S engraved on the edge of an opening, H , in its side. B is the specimen, and C is a rod of fused silica with a second scale S' engraved on it.

The thermal expansion of fused silica is much less than that of most solids, over a given temperature range, and has been accurately measured by a method depending on optical interference. When the apparatus shown is placed in a high-temperature bath, the rod C rises by an amount equal to the difference in expansion of the specimen, and an equal length of fused silica. Its rise is measured by observing the displacement of the scale S' relative to S through a microscope.

Force set up when Expansion is Resisted

Consider a metal rod between two supports P , which we suppose are immovable (Fig. 11.4). Let l_1 be the distance between the supports, and θ_1 the temperature at which the rod just fits between them. If the rod is heated to θ_2 , it will try to expand to a greater length l_2 , but will not be able to do so.

The value of l_2 would be

$$l_2 = l_1 \{1 + \lambda(\theta_2 - \theta_1)\},$$

and the expansion would be

$$e = l_2 - l_1 = l_1 \lambda (\theta_2 - \theta_1),$$

where λ is the mean linear expansivity of the rod.

The force which opposes the expansion is the force which would compress a bar of natural length l_2 by the amount e . Its magnitude F depends on the cross-section of the bar, A , and the Young's modulus of its material, E :

$$\frac{F}{A} = E \frac{e}{l_2}.$$

To a very good approximation we may replace l_2 by l_1 , because their difference is small compared with either of them. Thus

$$\begin{aligned} \frac{F}{A} &= E \frac{e}{l_1} = E \frac{\lambda l_1 (\theta_2 - \theta_1)}{l_1} \\ &= E \lambda (\theta_2 - \theta_1), \\ \therefore F &= EA \lambda (\theta_2 - \theta_1). \end{aligned}$$

For steel, $\lambda = 12 \times 10^{-6} \text{ K}^{-1}$ and $E = 2 \times 10^{11}$ newton per m^2 . If the temperature difference, $\theta_2 - \theta_1$, is 100°C , then, for a cross-sectional area S of 4 cm^2 or $4 \times 10^{-4} \text{ m}^2$,

$$\begin{aligned} F &= 2 \times 10^{11} \times 12 \times 10^{-6} \times 4 \times 10^{-4} \times 100 \text{ newton} \\ &= 9.6 \times 10^4 \text{ newton.} \end{aligned}$$

On converting to kgf we find F is nearly 10000 kgf .

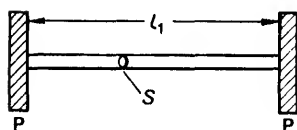


FIG. 11.4. Force in bar.

Expansion of a Measuring Scale

A scale, such as a metre rule, expands with rise in temperature; its readings, therefore, are correct at one temperature, θ_1 say. When the temperature of the scale is greater than θ_1 , the distance between any two of its divisions increases, and its reading is therefore too low (Fig. 11.5); when the scale is below θ_1 , its reading is too high. Let us suppose that, at θ_1 , the distance between any two points P, Q, on the scale is l_1 cm. At θ_2 it is

$$l_2 = l_1 \{1 + \lambda(\theta_2 - \theta_1)\},$$

where λ is the mean linear expansivity of the material of this scale. According to the divisions on the scale, however, the distance between P and Q will still be l_1 cm. Thus

$$\text{true distance at } \theta_2 = \text{scale value} \times \{1 + \lambda(\theta_2 - \theta_1)\} \quad (3)$$

If a sheet of material with a hole in it is warmed, it expands, and the hole expands with it. In Fig. 11.6, A represents a hole in a plate, and A' represents a plug, of the same material, that fits the hole. If A and A' are at the same temperature, then A' will fit A, whatever the value of that temperature; for we can always imagine A' to have just been cut out, without loss of material. It follows that the expansion of the hole A, in every direction, is the same as the expansion of the *solid* plug A'.

Differential Expansion

The difference in the expansions of different materials is used in practical arrangements discussed shortly. Fig. 11.7 (A) shows two rods AB, AB', of different metals, rigidly connected at A. If l_1, l'_1 are their lengths at a temperature θ_1 , their difference is

$$d_1 = BB' = l_1 - l'_1$$

If λ and λ' are the mean linear expansivities of the materials of the rods, then the lengths of the rods at θ_2 are

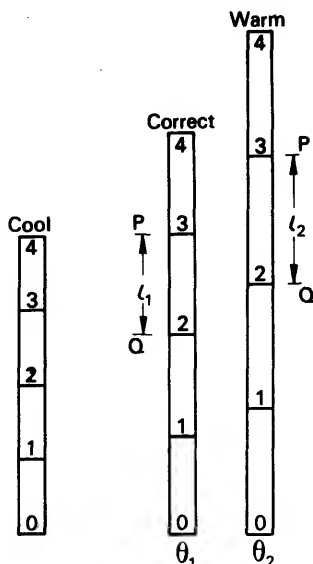


FIG. 11.5. Expansion of a measuring scale.

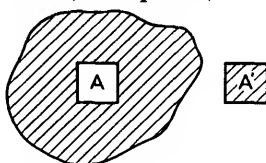


FIG. 11.6. Expansion of a hole.

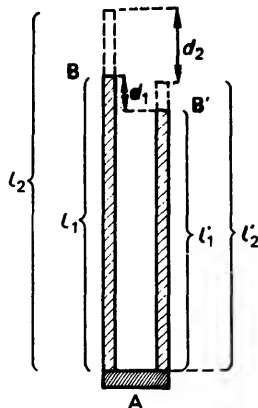


FIG. 11.7A. Differential expansion.

$$l_2 = l_1 \{1 + \lambda(\theta_2 - \theta_1)\},$$

$$l'_2 = l'_1 \{1 + \lambda'(\theta_2 - \theta_1)\}.$$

The distance between their ends is now

$$d_2 = l_2 - l'_2 = l_1 - l'_1 + (l_1\lambda - l'_1\lambda')(\theta_2 - \theta_1),$$

or
$$d_2 = d_1 + (l_1\lambda - l'_1\lambda')(\theta_2 - \theta_1) \quad (4)$$

By a suitable choice of lengths and materials, the distance BB' can be made to vary with temperature in any one of the following ways:

(1) The bar AB' is made of invar, a nickel-steel whose linear expansivity is very small (p. 270). The point B' then does not move with changes in temperature. In equation (4) we neglect λ' , and find:

$$d_2 = d_1 + l_1\lambda(\theta_2 - \theta_1).$$

Thus the short distance BB' expands by the same amount as the long

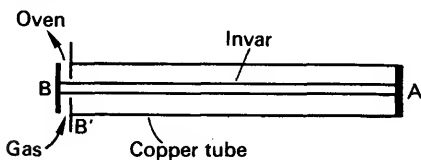


FIG. 11.7B. Thermostat principle.

distance AB . Consequently, the relative expansion of BB' with temperature is much greater than that of a bar of length d_1 .

(2) AB is made of invar, so that λ is negligible. The point B then does not move, and

the distance BB' *shrinks* rapidly as the temperature rises. This principle is used in the thermostats used to maintain gas ovens at constant temperatures (Fig. 11.7 (B)).

(3) The lengths and materials are chosen so that

$$\frac{l_1}{l'_1} = \frac{\lambda'}{\lambda},$$

or

$$l_1\lambda = l'_1\lambda'.$$

Then, by equation (4), d_1 or BB' does not change with temperature. This principle is used in compensating clock pendula for temperature changes (p. 276).

Bimetal Strip

Fig. 11.8 (a) shows two strips of different metals, welded together along B , called a bimetal strip. The metal M_1 has a greater linear expansivity than the metal M_2 . Therefore, when the strip is heated, M_1 will expand more than M_2 , and the strip will curl with M_1 on the outside. The reverse is true when the strip is cooled, as M_1 then shrinks more than M_2 (see Fig. 11.8(a)).

Bi-metal strips are used in electrical thermostats for ovens, irons, laboratories, etc. The strip carries a contact, K in Fig. 11.8 (b), which presses against another contact K' on the end of an adjusting screw S . When the strip warms, it tends to curl away from K' ; the temperature at which the contact is broken can be set by turning the screw. When

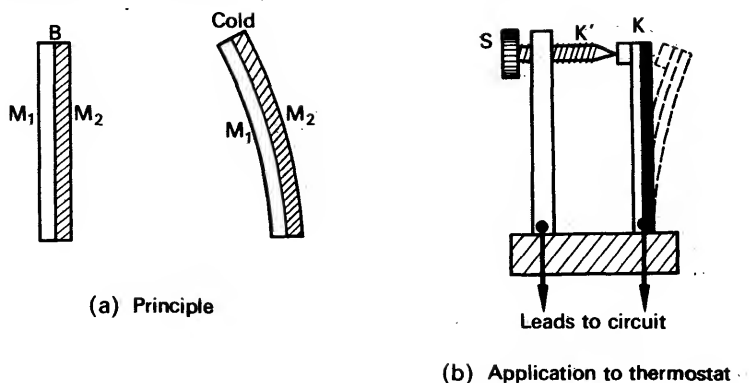


FIG. 11.8. Principle of gas thermostat:

the contacts open, they switch off the heating current. If the heating current is too great to be controlled by the contacts KK' , it is switched off by a relay, which is controlled by the contacts on the bimetal.

Let us now estimate the deflection of a heated bimetal strip. We suppose that the component strips have the same thickness d , and the same length l at a temperature θ_1 (Fig. 11.9 (a)). When they are heated, they are distorted, but to a first approximation we may assume that the mid-line of each, shown dotted, has the length which it would naturally have (Fig. 11.9 (b)). The difference in length of the mid-lines, p , is then the difference in their expansions. At a temperature θ_2 ,

$$\begin{aligned} p &= l\lambda(\theta_2 - \theta_1) - l\lambda'(\theta_2 - \theta_1) \\ &= l(\lambda - \lambda')(\theta_2 - \theta_1). \end{aligned}$$

The difference is taken up by the curvature of the strip. If α is the angle through which it bends, then, from the figure,

$$\alpha = \frac{p}{d} = \frac{l}{d}(\lambda - \lambda')(\theta_2 - \theta_1).$$

(The expansion of d is negligible, to a very good approximation.)

To find the radius r of the arc formed by the strip, we assume that the length of the arc is l ; this we may do because the expansions in length are all small. Then

$$\alpha = \frac{l}{r};$$

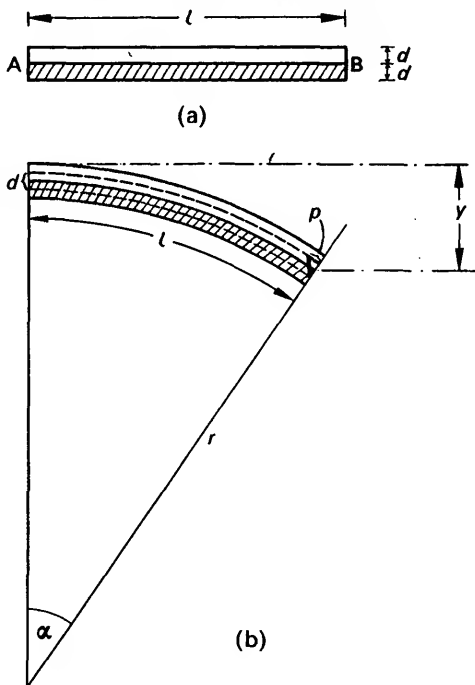


FIG. 11.9. Expansion of bimetal strip (exaggerated).

and, by the above equation,

$$\frac{l}{r} = \frac{l}{d}(\lambda - \lambda')(\theta_2 - \theta_1),$$

whence

$$r = \frac{d}{(\lambda - \lambda')(\theta_2 - \theta_1)}.$$

The deflection y of the end of the strip is given by the approximate equation

$$2ry = l^2,$$

from the geometry of the circle.

Thus

$$y = \frac{l^2}{2r} = \frac{l^2(\lambda - \lambda')(\theta_2 - \theta_1)}{2d}.$$

Temperature Compensation

The rate of a clock varies considerably with temperature, unless arrangements are made to prevent its doing so. If the clock is governed by a pendulum, the length of the pendulum increases with temperature, its period therefore also increases, and the clock loses. A clock governed by a balance-wheel and hair-spring also loses as the temperature rises. For, as the temperature rises, the spring becomes less stiff, and the period of the balance-wheel increases. Also the spokes of the wheel expand a little, increasing the moment of inertia of the wheel, and thus further increasing its period.

A balance-wheel clock may be compensated against temperature changes by making the circumference of the wheel in the form of two or three bimetal strips, as shown in Fig. 11.10. The strips

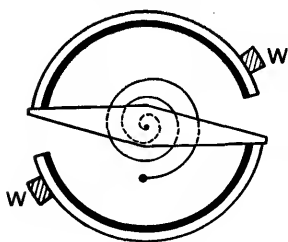


FIG. 11.10. Bimetal balance-wheel.

carry small weights W , to give the wheel the necessary moment of inertia. As the temperature rises, the strips curl inwards, and bring the weights nearer to the axle; thus the moment of inertia of the wheel decreases. In a correctly designed timing-system, the decrease in moment of inertia just offsets the decrease in stiffness of the spring, and then the period of the balance-wheel does not change with temperature.

Many modern watches are not compensated. Their balance-wheels are made of invar, which, as we have seen, has a very small coefficient of expansion. Another nickel-steel, of slightly different composition, is used for their hair-springs. This alloy changes its elasticity very little with temperature, and is called *elinvar*. The combination of invar balance-wheel and elinvar hair-spring gives a rate which is nearly enough independent of temperature for everyday purposes.

Pendula

To compensate a pendulum clock against changes of temperature, the pendulum must be so made that its effective length remains con-

stant. Fig. 11.11 (a) shows one way of doing this, in the so-called grid-iron pendulum (Harrison, 1761). Brass and steel rods are arranged so that the expansion of the brass rods raises the bob B of the pendulum, while the expansion of the steel rods lowers it. As explained on p. 274, the expansions can be made to cancel if

$$\frac{l_B}{l_S} = \frac{\lambda_S}{\lambda_B},$$

here l_B and l_S are the total lengths of brass and steel respectively, and λ_B , λ_S are their linear expansivities.

Fig. 11.11 (b) shows the same principle applied to a pendulum with a wooden rod and a cylindrical metal bob. To a first approximation the effective length of the pendulum, l , is the distance from its support to the centre of gravity, G, of the bob. The condition for constant period is then

$$\frac{\text{length of rod}}{\frac{1}{2} \text{ length of bob}} = \frac{\lambda_{\text{wood}}}{\lambda_{\text{metal}}}.$$

The bob may be made of lead or zinc, either of which is much more expansible than is wood, along its grain.

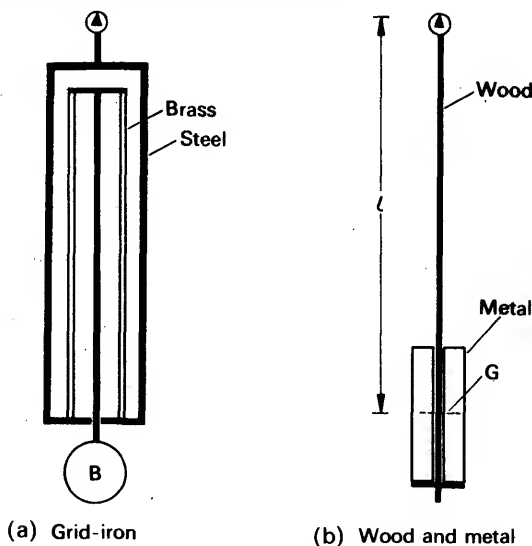


FIG. 11.11. Compensated pendula.

Metal-Glass Seals

In radio valves and many other pieces of physical apparatus, it is necessary to seal metal cones into glass tubes, with a vacuum-tight joint. The seal must be made at about 400°C , when the glass is soft; as it cools to room temperature, the glass will crack unless the glass and metal contract at the same rate. This condition requires that the metal and the glass have the same linear expansivity at every temperature between room temperature and the melting-point of

glass. It is satisfied nearly enough by platinum and soda glass (mean $\lambda = 9$ and 8.5×10^{-6} per deg C, respectively), and by tungsten and some types of hard glass similar to pyrex (mean linear expansivity $\lambda = 3-4 \times 10^{-6}$ per deg C).

Modern seals through soft glass are not made with platinum, but with a wire of nickel-iron alloy, which has about the same linear coefficient as the glass. The wire has a thin coating of copper, which adheres to glass more firmly than the alloy. Also, being soft, the copper takes up small differences in expansion between the alloy and the glass.

In transmitting valves, and large vacuum plants, glass and metal tubes several centimetres in diameter must be joined end-to-end. The metal tubes are made of copper, chamfered to a fine taper at the end where the joint is to be made. The glass is sealed on to the edge of the chamfer; the copper there is thin enough to distort, with the difference in contraction, without cracking the glass.

Superficial Expansion

The increase in area of a body with temperature change is called the *superficial expansion* of the body. A rectangular plate, of sides a , b , at a given temperature, has an area

$$S_1 = ab.$$

If its temperature is increased by θ its sides become $a(1 + \lambda\theta)$, $b(1 + \lambda\theta)$ where λ is its mean linear expansivity. Thus its area becomes

$$\begin{aligned} S_2 &= a(1 + \lambda\theta) b(1 + \lambda\theta) \\ &= S_1(1 + \lambda\theta)^2 \\ &= S_1(1 + 2\lambda\theta + \lambda^2\theta^2). \end{aligned}$$

In this expression, the term $\lambda^2\theta^2$ is small compared with $2\lambda\theta$; if λ is of the order of 10^{-5} , and θ of the order of 100, then $\lambda\theta \approx 10^{-3}$, and $\lambda^2\theta^2 \approx 10^{-6}$. Therefore we may neglect $\lambda^2\theta^2$ and write

$$S_2 = S_1(1 + 2\lambda\theta).$$

The superficial expansivity of the material of the plate is defined as

$$\frac{\text{increase of area}}{\text{original area} \times \text{temp. rise}} = \frac{S_2 - S_1}{S_1\theta}.$$

Its value is hence equal to 2λ , twice the linear expansivity. A hole in a plate changes its area by the same amount as would a plug that fitted the hole.

Cubic Expansivity

Cubic expansivity is expansion in volume. Consider a rectangular block of sides a , b , c , and therefore of volume

$$V_1 = abc.$$

If the block is raised in temperature by θ its sides expand, and its volume becomes

$$\begin{aligned} V_2 &= a(1 + \lambda\theta) b(1 + \lambda\theta) c(1 + \lambda\theta) = abc(1 + \lambda\theta)^3 \\ &= abc(1 + 3\lambda\theta + 3\lambda^2\theta^2 + \lambda^3\theta^3). \end{aligned}$$

Since $\lambda^2\theta^2$ and $\lambda^3\theta^3$ are small compared with $\lambda\theta$, we may in practice neglect them. We then have

$$V_2 = abc(1 + 3\lambda\theta) = V_1(1 + 3\lambda\theta).$$

The cubic expansivity of the solid is defined as

$$\begin{aligned} \gamma &= \frac{\text{increase in volume}}{\text{original volume} \times \text{temp. rise}} \\ &= \frac{V_2 - V_1}{V_1\theta} \\ &= 3\lambda, \text{ to a very good approximation.} \end{aligned}$$

Thus the cubic expansivity is *three* times the linear expansivity.

By imagining a block cut out of a larger block, we can see that the cubical expansion of a hollow vessel is the same as that of a solid plug which would fit into it.

LIQUIDS

Cubic Expansivity

The temperature of a liquid determines its volume, but its vessel determines its shape. The only expansivity which we can define for a liquid is therefore its cubic expansivity, γ . Most liquids, like most solids, do not expand uniformly, and γ is not constant over a wide range of temperature. Over a given range θ_1 to θ_2 , the *mean coefficient* γ is defined as

$$\gamma = \frac{V_2 - V_1}{V_1(\theta_2 - \theta_1)},$$

where V_1 and V_2 are the volumes of a given mass of liquid at the temperatures θ_1 and θ_2 .

MEAN EXPANSIVITIES OF LIQUIDS
(Near room temperature)

Liquid	γ (K ⁻¹) $\times 10^{-4}$	Liquid	γ (K ⁻¹) $\times 10^{-4}$
Alcohol (methyl)	12.2	Water: 5–10°C	0.53
Alcohol (ethyl)	11.0	10–20°C	1.50
Aniline	8.5	20–40°C	3.02
Ether (ethyl)	16.3	40–60°C	4.58
Glycerine	5.3	60–80°C	5.87
Olive oil	7.0	Mercury: 0–30°C	1.81
Paraffin oil	9.0	0–100°C	1.82
Toluene	10.9	0–300°C	1.87

shown in Fig. 11.13. It consists of a glass tube ABCD, a foot or two high, containing mercury, and surrounded by glass jackets XY. The jacket X contains ice-water, and steam is passed through the jacket Y.

For the mercury to be in equilibrium, its hydrostatic pressure at B must equal its hydrostatic pressure at C. Let h_0 be the height of the mercury in the limb at 0°C and ρ_0 its density; and let h and ρ be the corresponding quantities at the temperature θ of the steam.

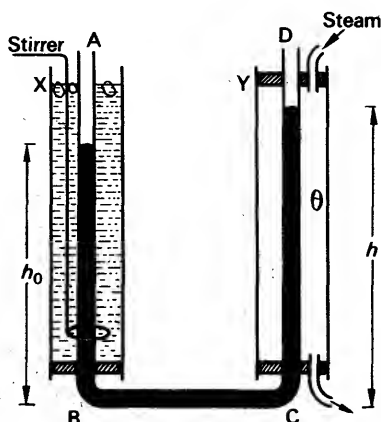


FIG. 11.13. Apparatus for true (absolute) expansivity of mercury.

Then $g\rho_0 h_0 = g\rho h$,

where g is the acceleration of gravity.

Hence
$$\frac{\rho}{\rho_0} = \frac{h_0}{h}.$$

But, by equation (5),

$$\begin{aligned}\frac{\rho}{\rho_0} &= \frac{1}{1 + \gamma\theta} \\ \therefore \frac{h_0}{h} &= \frac{1}{1 + \gamma\theta}\end{aligned}$$

or $h_0 + h_0\gamma\theta = h$.

$$\therefore \gamma = \frac{h - h_0}{h_0\theta}.$$

The height $h - h_0$ is measured with a cathetometer (a travelling telescope on a vertical column).

This simple apparatus is inaccurate because:

- (i) the expansion of CD throws BC out of the horizontal;
- (ii) the wide separation of A and D makes the measurement of $(h - h_0)$ inaccurate;
- (iii) surface tension causes a difference of pressure across each free surface of mercury; and these do not cancel one another, because the surface tensions are different at the temperatures of the hot and cold columns.

Regnault got round these difficulties with the apparatus shown, somewhat simplified, in Fig. 11.14. The points B and G are fixed at the same horizontal level, the join DE is made of flexible iron tubing, and

the difference in height between its ends, h_2 , is measured. The parts AB, GH are at room temperature θ_1 ; and to a fair approximation, the average temperature of DE is also θ_1 . Suppose θ is the steam temperature.

If the density of mercury at θ_1 , θ is ρ_1 , ρ , respectively then equating the pressures on both sides at the horizontal level of E, we have

$$g\rho_1 h_1 + g\rho_0 h_0 + g\rho_1 h_2 = g\rho h + g\rho_1 h_3.$$

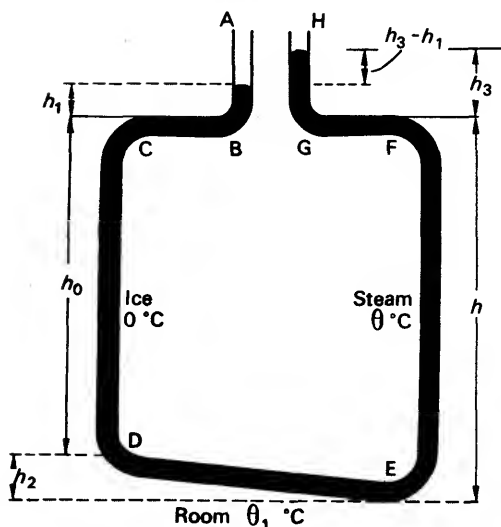


FIG. 11.14. Modified apparatus for true expansivity

Therefore*

$$\frac{\rho_0 h_1}{1 + \gamma_1 \theta_1} + \rho_0 h_0 + \frac{\rho_0 h_2}{1 + \gamma_1 \theta_1} = \frac{\rho_0 h}{1 + \gamma \theta} + \frac{\rho_0 h_3}{1 + \gamma_1 \theta_1}. \quad (6)$$

The uncertainty in the temperature of DE is not important, since the height h_2 is very small. Equation (6) gives

$$h_0 + \frac{h_2}{1 + \gamma_1 \theta_1} = \frac{h}{1 + \gamma \theta} + \frac{h_3 - h_1}{1 + \gamma_1 \theta_1}. \quad (7)$$

Equation (7) can be solved for γ : the quantities which need to be known accurately are h_0 , h , and the difference $h_3 - h_1$. This difference can be measured accurately, because AB and GH are close together; and because they are at the same temperature there is no error due to surface tension. The heights h_0 and h are 1 or 2 metres, and so are easy to measure accurately.

Callender and Moss used six pairs of hot and cold columns, each 2 metres long, to increase the difference in level of the liquid. In this way they avoided the complication due to density change of the liquid under high pressure. All the hot columns were beside one another in a hot oil bath kept at a constant high temperature, while the cold columns were similarly placed in a bath of melting ice. Platinum resistance thermometers were used to measure the temperatures, and a cathetometer to measure the heights and difference in level.

* Strictly, equation (6) should be written

$$\frac{\rho_0 h_1}{1 + \gamma_1 \theta_1} + \rho_0 h_0 + \frac{\rho_0 h_2}{1 + \gamma_1 \theta_1} = \frac{\rho_0 h}{1 + \gamma \theta} + \frac{\rho_0 h_3}{1 + \gamma_1 \theta_1},$$

where γ_1 is the mean coefficient between 0°C and θ_1 , and γ is that between 0°C and θ . This equation can be solved for γ by successive approximations. (See Roberts-Miller, *Heat and Thermodynamics*, Chapter X. Blackie.)

Apparent Expansion: Weight Thermometer Method

The method of balancing columns for the absolute expansivity of a liquid is slow and awkward; it has only been applied to mercury. Routine measurements are more conveniently made by measuring the apparent expansion; from this, as we shall see, the absolute expansion can be calculated.

A *weight thermometer* is a bulb of fused quartz fitted with a fine stem (Fig. 11.15). It is filled with liquid at a low temperature, and then warmed; the liquid expands, and from the mass which flows out the apparent expansion of the liquid can be found. The weight thermometer is filled by warming it to expel air, and then dipping the stem into the liquid. The process has to be repeated many times; and for accurate work the liquid in the thermometer should be boiled at intervals during the filling, to expel dissolved air. In a laboratory experiment, a glass density bottle may be used, filled in the usual way.

The weight thermometer must be filled at a temperature slightly below the lower limit, θ_1 , of the range over which the expansion is to be measured. It is then kept in a bath at θ_1 , until no more liquid flows out. Next it is weighed, and from its known mass when empty the mass of liquid in it is found. Let this be m_1 . The weight thermometer is then placed in a bath at the higher temperature of the range, θ_2 , and the mass remaining in it, m_2 , is found by weighing.

If V_1 and V_2 are the volumes of the weight thermometer at θ_1 and θ_2 , then

$$V_2 = V_1 \{1 + 3\lambda(\theta_2 - \theta_1)\},$$

where λ is the linear expansivity of quartz, the material of the weight thermometer. Now, if ρ denotes the density of the liquid, whose cubic expansivity is γ , we have

$$m_1 = V_1 \rho_1,$$

$$m_2 = V_2 \rho_2,$$

and

$$\frac{\rho_2}{\rho_1} = \frac{1}{1 + \gamma(\theta_2 - \theta_1)}.$$

Therefore

$$\begin{aligned} \frac{m_2}{m_1} &= \frac{V_2 \rho_2}{V_1 \rho_1} \\ &= \frac{1 + 3\lambda(\theta_2 - \theta_1)}{1 + \gamma(\theta_2 - \theta_1)}. \end{aligned}$$

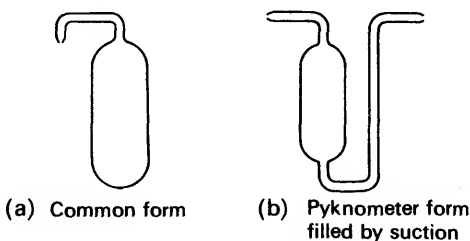


FIG. 11.15. Weight thermometers.

Hence $m_2 + m_2\gamma(\theta_2 - \theta_1) = m_1 + 3\lambda m_1(\theta_2 - \theta_1)$

or $m_2\gamma(\theta_2 - \theta_1) = m_1 - m_2 + 3\lambda m_1(\theta_2 - \theta_1)$

and
$$\gamma = \frac{m_1 - m_2}{m_2(\theta_2 - \theta_1)} + 3\lambda \frac{m_1}{m_2} \quad (11)$$

If we ignore the expansion of the solid, we obtain the apparent expansivity of the liquid, γ_a , relative to the solid.

Thus
$$\gamma_a = \frac{m_1 - m_2}{m_2(\theta_2 - \theta_1)} \quad (12)$$

The expression for the apparent expansivity may be expressed in words:

$$\gamma_a = \frac{\text{mass expelled}}{\text{mass left in } \times \text{ temp. rise}}$$

Attention is often drawn to the fact that the mass in the denominator is the mass at the higher temperature, and not the lower.

From equation (11),

$$\gamma = \gamma_a + \frac{m_1}{m_2} \gamma_g \quad (13a)$$

where $\gamma_g = 3\lambda =$ the cubical coefficient of the solid. Since m_1 and m_2 are nearly equal, then, to a good approximation,

$$\gamma = \gamma_a + \gamma_g \quad (13b)$$

Thus after the apparent coefficient has been calculated, the absolute coefficient is given to a good approximation by

$$\text{absolute expansivity} = \text{apparent expansivity} + \text{cubic expansivity of vessel.}$$

Coefficient of Weight-thermometer Material

The cubic expansivity of a glass weight-thermometer may not be accurately known, because glasses differ considerably in their physical properties. Also the expansivity may be changed when the glass is heated in the blowing of the bulb. The expansivity can conveniently be measured, however, by using the weight thermometer to find the apparent value for mercury; and then subtracting the value found from the known value of the absolute expansivity of mercury.

Expansion of a Powder

The weight thermometer can be used to find the cubic expansivity of a granular or powdery solid, such as sand. The procedure is the same as in finding the relative density of the solid, but is gone through at two known temperatures. From the change in relative density, and a knowledge of the expansions of the liquid and the weight thermometer, the change in absolute density of the powder can be found, and hence its cubic expansivity.

The Dilatometer

A dilatometer is an instrument for rapidly—but roughly—measuring the expansion of a liquid. It consists of a glass bulb B, with a graduated stem S (Fig. 11.16). The volume V_b of the bulb, up to the zero of S, is known, and S is graduated in cubic millimetres or other small units. The volume of the bulb, and the value of one scale division, vary with temperature; the dilatometer therefore measures apparent expansion.



FIG. 11.16. Dilatometer.



FIG. 11.17. Compensated dilatometer.

The dilatometer is filled with the liquid under test to a point just above the zero of the stem, at a temperature θ_1 . The volume V_1 of the liquid is found by adding V_b to the stem-reading v_1 . Next the dilatometer is warmed to θ_2 , and the liquid rises to v_2 . Then $(v_2 - v_1)$ is the apparent expansion of the liquid, and hence

$$\gamma_a = \frac{v_2 - v_1}{V_1(\theta_2 - \theta_1)}$$

If γ_g is the cubical coefficient of the glass, then

$$\gamma = \gamma_a + \gamma_g.$$

For demonstration work, a dilatometer can be compensated so that it shows roughly the true expansion of a liquid. Mercury is introduced into the bulb, until it occupies $1/7$ th of the bulb's volume (Fig. 11.17). The expansion of the mercury is then about equal to the expansion of the glass, so that the free volume in the bulb is roughly constant. The cubic expansivities of mercury and glass are given respectively by

$$\gamma_{Hg} = 18.1 \times 10^{-5} \text{ K}^{-1}$$

and $\gamma_g = 3\lambda_g = 3 \times 8.5 \times 10^{-6} = 2.55 \times 10^{-5} \text{ K}^{-1}.$

Thus
$$\frac{\gamma_{Hg}}{\gamma_g} = \frac{18.1}{2.55} = 7.1.$$

Thus the expansion of the mercury offsets that of the glass, within about $1\frac{1}{2}$ per cent.

The space above the mercury, whose volume is constant, is filled

with the liquid to be examined. When the bulb is warmed, the movement of the liquid up the stem shows the liquid's true expansion. This device may be used to show the anomalous expansion of water (p. 288).

Correction of the Barometer

The hydrostatic pressure of a column of mercury, such as that in a barometer, depends on its density as well as its height. When we speak of a pressure of 760 mm mercury, therefore, we must specify the temperature of the mercury; we choose 0°C . In practice barometers are generally warmer than that, and their readings must therefore be

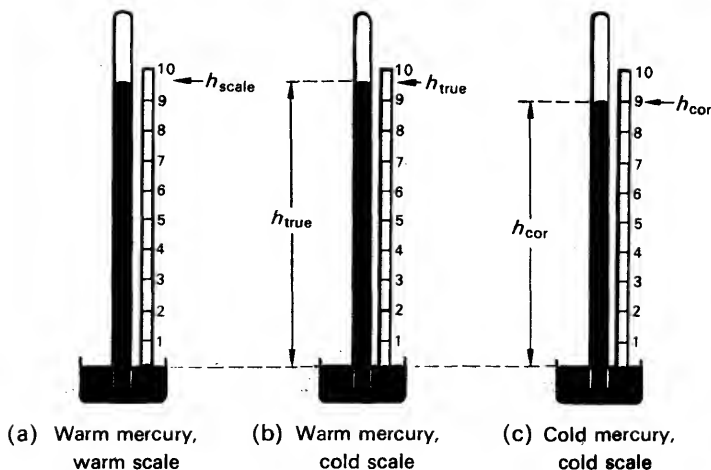


FIG. 11.18. Reduction of barometer height to 0°C .

reduced to what they would be at 0°C . Also we must allow for the expansion of the scale with which the height is measured.

The scale of a barometer may be calibrated at 0°C . At any higher temperature, θ , the height which it indicates, h_{scale} , is less than the true height, h_{true} , of the mercury meniscus above the free surface in the reservoir (Fig. 11.18 (a) (b)). The true height is given by equation (3) of p. 273.

$$h_{\text{true}} = h_{\text{scale}} (1 + \lambda\theta) \quad (14)$$

where λ is the linear expansivity of the scale. If ρ_θ is the density of mercury at θ , then the pressure of the atmosphere is

$$p = g\rho_\theta h_{\text{true}}.$$

The height of a mercury column at 0°C which would exert the same pressure is called the *corrected height*, $h_{\text{cor.}}$ (Fig. 11.18 (b) (c)). It is given by

$$p = g\rho_0 h_{\text{cor.}},$$

where ρ_0 is the density of mercury at 0°C .

Therefore

$$\rho_{\theta} h_{\text{true}} = \rho_0 h_{\text{cor.}}$$

If γ is the coefficient of expansion of mercury, then

$$\rho_{\theta} = \frac{\rho_0}{1 + \gamma\theta}.$$

Hence

$$\frac{\rho_0}{1 + \gamma\theta} h_{\text{true}} = \rho_0 h_{\text{cor.}}$$

and

$$h_{\text{cor.}} = \frac{h_{\text{true}}}{1 + \gamma\theta}.$$

Therefore, by equation (13),

$$h_{\text{cor.}} = \frac{h_{\text{scale}} (1 + \lambda\theta)}{1 + \gamma\theta}.$$

Let us write this as

$$h_{\text{cor.}} = h_{\text{scale}} (1 + \lambda\theta)(1 + \gamma\theta)^{-1}.$$

Then, if we ignore $\gamma^2\theta^2$ and higher terms, we may write

$$h_{\text{cor.}} = h_{\text{scale}} (1 + \lambda\theta)(1 - \gamma\theta).$$

Hence

$$h_{\text{cor.}} = h_{\text{scale}} (1 + \lambda\theta - \gamma\theta + \gamma\lambda\theta^2),$$

and ignoring $\gamma\lambda\theta^2$ we find

$$h_{\text{cor.}} = h_{\text{scale}} \{1 + (\lambda - \gamma)\theta\}.$$

The coefficient γ is greater than λ , and the corrected height is less than the scale height. It is convenient therefore, to write

$$h_{\text{cor.}} = h_{\text{scale}} \{1 - (\gamma - \lambda)\theta\}.$$

The reader should notice that the correction depends on the difference between the cubic expansivity of the mercury, and the linear expansivity of the scale; as in Dulong and Petit's experiment, there is no question of apparent expansion.

The Anomalous Expansion of Water

If we nearly fill a tall jar with water, and float lumps of ice on it, the water at the base of the jar does not cool below 4°C , although the water at the top soon reaches 0°C . At 4°C water has its greatest density; as it cools to this temperature it sinks to the bottom. When the water at the top cools below 4°C , it becomes less dense than the water below, and stays on the top. Convection ceases, and the water near the bottom of the jar can lose heat only by conduction. Since water is a bad conductor, the loss by conduction is extremely small, and in practice the water at the bottom does not cool below 4°C . The same happens in a pond, cooled at the top by cold air. Ice forms at the surface, but a little below it the water remains at 4°C , and life in the pond survives. It

could not if the water contracted in volume continuously to 0°C ; for then convection would always carry the coldest water to the bottom, and the pond would freeze solid. In fact, lakes and rivers, unless they are extremely shallow, never do freeze solid; even in arctic climates, they take only a crust of ice.

Fig. 11.19 shows how the volume of 1 g of water varies with temperature. The decrease from 0°C to 4°C is called *anomalous expansion*; it can be shown with a compensated dilatometer (p. 285). Water also has the unusual property of expanding when it freezes, hence burst

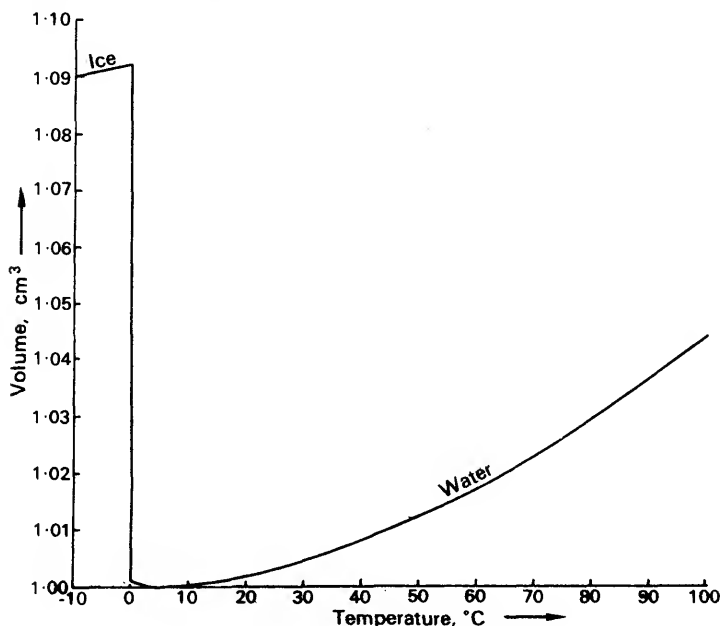


FIG. 11.19. Volume variation of 1 g of water.

pipes. As the figure shows, the expansion is about 9 per cent. If ice were not less dense than water, it would sink, and so, despite the anomalous expansion of water, lakes and rivers would freeze solid in winter.

The following table gives the densities of ice and water at various temperatures. Accurate experiments show that the temperature of maximum density is 3.98°C .

DENSITY OF ICE AND WATER

Temperature °C .	0	2	4	6	8	
ρ_{water} kg m ⁻³ .	999.84	999.94	999.97	999.94	999.85	
ρ_{ice} kg m ⁻³ .	916.0 (volume of 1 g ice at 0°C = 1.092 cm ³)					
Temperature °C .	10	20	40	60	80	100
ρ_{water} kg m ⁻³ .	999.70	998.20	992.2	983.2	971.8	958.4

The temperature of maximum density was first measured in 1804, by Hope, whose apparatus is shown in Fig. 11.20. Both vessels are made of metal; the inner one contains water, and the outer one contains a mixture of ice and salt, which has a temperature below 0°C . After several hours, the water at the top cools to 0°C —it may even freeze—but the water at the bottom does not fall below 4°C .

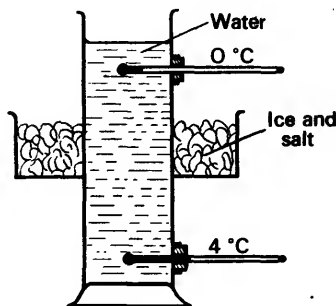


FIG. 11.20. Hope's experiment.

The contraction of water from 0 to 4°C is explained by supposing that the molecules form clusters, such as H_4O_2 , H_6O_3 . At first, the contraction due to the formation of these more than offsets the expansion due to the rise of temperature; but above 4°C the expansion prevails. The metal antimony behaves like water: it expands on freezing, and contracts at first when warmed above its melting-point. Because it expands on freezing, it makes sharp castings, and is used as a constituent of type-metal.

EXAMPLES

1. Describe how you would use a specific gravity bottle to find the coefficient of expansion of paraffin oil relative to glass between 0° and 50°C .

A specific gravity bottle contains 44.25 g of a liquid at 0°C and 42.02 g at 50°C . Assuming that the linear expansivity of the glass is 0.00001 K^{-1} , (a) compare the densities of the liquid at 0° and 50°C , (b) deduce the real expansivity of the liquid. Prove any formula employed. (N.)

First part. See text.

Second part. The apparent expansivity of the liquid, γ_a , is given by

$$\gamma_a = \frac{\text{mass expelled}}{\text{mass left} \times \text{temp. rise}} = \frac{44.25 - 42.02}{42.02 \times 50}.$$

$$\therefore \gamma_a = \frac{2.23}{42.02 \times 50} = 0.00106\text{ K}^{-1}.$$

Now cubic expansivity, γ_g , of glass $= 3 \times 0.00001 = 0.00003\text{ K}^{-1}$.

$$\therefore \text{real coefficient, } \gamma_r = \gamma_a + \gamma_g = 0.00106 + 0.00003 = 0.00109\text{ K}^{-1}.$$

Also, if ρ_0, ρ_t are the densities at 0°C , 50°C respectively,

$$\begin{aligned} \frac{\rho_0}{\rho_t} &= 1 + \gamma t \\ &= 1 + 0.00109 \times 50 \\ &= 1.055. \end{aligned}$$

2. Define the linear expansivity of a solid, and show how it is related to the cubic expansivity. Describe an accurate method of determining the coefficient of expansion of a solid.

In order to make connexion to the carbon anode of a transmitting valve it is

required to thread a copper wire (2 mm in diameter at 20°C) through a smaller hole in the carbon block. If the process can just be carried out by immersing both specimens in dry ice (solid CO_2) at -80°C and the coefficients of linear expansion of copper and carbon are 17×10^{-6} and $5 \times 10^{-6} \text{ K}^{-1}$, calculate the size of the hole at 20°C . (*O. & C.*)

First part. The linear expansivity is the increase in length per unit length per degree rise in temperature. It is one-third of the cubic expansivity, proved on p. 279. An accurate method of measuring the linear expansivity is the comparator method, described on p. 270.

Second part. Suppose r cm is the diameter of the hole at 20°C . Then, from the formula

$$l_2 = l_1[1 + \lambda(\theta_2 - \theta_1)], \text{ where } (\theta_2 - \theta_1) = -80 - 20 = -100^{\circ}\text{C},$$

$$\text{diameter at } -80^{\circ}\text{C} = r(1 - 5 \times 10^{-6} \times 100) = 0.9995 r \text{ cm.}$$

Similarly, the diameter of the copper wire at -80°C

$$= 0.2(1 - 17 \times 10^{-6} \times 100) = 0.19966 \text{ cm.}$$

$$\therefore 0.9995 r = 0.19966$$

$$\therefore r = \frac{0.19966}{0.9995} = 0.19976 \text{ cm} = 1.998 \text{ mm.}$$

3. Describe in detail how the cubic expansivity of a liquid may be determined by the weight thermometer method.

The height of the mercury column in a barometer provided with a brass scale correct at 0°C is observed to be 749.0 mm on an occasion when the temperature is 15°C . Find (a) the true height of the column at 15°C , (b) the height of a column of mercury at 0°C which would exert an equal pressure. Assume that the cubic expansivities of brass and of mercury are respectively 0.000054 and 0.000181 K^{-1} . (*N.*)

First part. The weight thermometer method gives the apparent expansivity of the liquid, which is calculated from:

$$\frac{\text{mass expelled}}{\text{mass left} \times \text{temperature rise}}$$

The true expansivity is obtained by adding the cubic expansivity of the container's material to the apparent expansivity.

Second part. (a) The linear expansivity of brass, λ ,

$$= \frac{1}{3} \times \text{volume expansivity}$$

$$= \frac{1}{3} \times 0.000054 = 0.000018 \text{ K}^{-1}.$$

$$\therefore \text{true height, } h_0 = 749.0(1 + \lambda\theta) = 749.0(1 + 0.000018 \times 15)$$

$$= 749.2 \text{ mm.}$$

(b) Suppose h_0 is the required height at 0°C . Then, since pressure = $h\rho g$,

$$h_0\rho_0g = h_\theta\rho_\theta g$$

$$\therefore h_0 = h_\theta \frac{\rho_\theta}{\rho_0}.$$

But
$$\frac{\rho_\theta}{\rho_0} = \frac{1}{1+\gamma\theta} = \frac{1}{1+0.000181 \times 15} = \frac{1}{1.0027}$$

$$\therefore h_0 = 749.0 \times \frac{1.00027}{1.0027}$$

$$= 749 [1 - (0.0027 - 0.00027)]$$

$$= 747.2 \text{ mm.}$$

EXERCISES 11

1. Describe and explain how the absolute expansivity of a liquid may be determined without a previous knowledge of any other expansivity.

Aniline is a liquid which does not mix with water, and when a small quantity of it is poured into a beaker of water at 20°C it sinks to the bottom, the densities of the two liquids at 20°C being 1021 and 998 kg m⁻³ respectively. To what temperature must the beaker and its contents be uniformly heated so that the aniline will form a globule which just floats in the water? (The mean absolute expansivity of aniline and water over the temperature range concerned are 0.00085 and 0.00045 K⁻¹, respectively.) (L.)

2. Define the *linear expansivity* of a solid, and describe a method by which it may be measured.

Show how the *superficial expansivity* can be derived from this value.

A 'thermal tap' used in certain apparatus consists of a silica rod which fits tightly inside an aluminium tube whose internal diameter is 8 mm at 0°C. When the temperature is raised, the fit is no longer exact. Calculate what change in temperature is necessary to produce a channel whose cross-section is equal to that of a tube of 1 mm internal diameter. [Linear expansivity of silica = 8×10^{-6} K⁻¹. Linear expansivity of aluminium = 26×10^{-6} K⁻¹.] (O. & C.)

3. Define the *cubic expansivity* of a liquid. Find an expression for the variation of the density of a liquid with temperature in terms of its expansivity.

Describe, without experimental details, how the cubic expansivity of a liquid may be determined by the use of balanced columns.

A certain Fortin barometer has its pointers, body and scales made from brass. When it is at 0°C it records a barometric pressure of 760 mm Hg. What will it read when its temperature is increased to 20°C if the pressure of the atmosphere remains unchanged?

[Cubic expansivity of mercury = 1.8×10^{-4} K⁻¹; linear expansivity of brass = 2×10^{-5} K⁻¹.] (O. & C.)

4. Describe an experiment to determine the absolute (real) expansivity of paraffin between 0°C and 50°C, using a weight thermometer made of glass of known cubic expansivity. Derive the formula used to calculate the result.

A glass capillary tube with a uniform bore contains a thread of mercury 100.0 cm long, the temperature being 0°C. When the temperature of the tube and mercury is raised to 100°C the thread is increased in length by 1.65 cm. If the mean coefficient of absolute expansion of mercury between 0°C and 100°C is 0.000182 K⁻¹, calculate a value for the linear expansivity of glass (N.)

5. Define the linear and cubic expansivity, and derive the relation between them for a particular substance. Describe, and give the theory of, a method for finding *directly* the absolute coefficient of expansion of a liquid. The bulb of a mercury-in-glass thermometer has a volume of 0.5 cm³ and the distance between progressive degree marks is 2 mm. If the linear expansivity of glass is 10⁻⁵ K⁻¹,

and the cubic expansivity of mercury is $1.8 \times 10^{-4} \text{ K}^{-1}$, find the cross-sectional area of the bore of the stem. (C.)

6. Define the terms 'apparent' and 'absolute' expansivity of a liquid, and show how the former is found by means of a weight thermometer. A litre flask, which is correctly calibrated at 4°C , is filled to the mark with water at 80°C . What is the weight of water in the flask? [Linear expansivity of the glass of the flask = $8.5 \times 10^{-6} \text{ K}^{-1}$; mean cubic expansivity of water = $5.0 \times 10^{-4} \text{ K}^{-1}$.] (O. & C.)

7. Describe how to measure the apparent expansivity of a liquid using a weight thermometer. Show how the result can be calculated from the observations.

A specific gravity bottle of volume 50.0 cm^3 at 0°C is filled with glycerine at 20°C . What mass of glycerine is contained in the bottle if the density of glycerine at 0°C is 1.26 g per cm^3 , and its real expansivity is $5.2 \times 10^{-4} \text{ K}^{-1}$? Assume that the linear expansivity of the glass is $8 \times 10^{-6} \text{ K}^{-1}$. (N.)

8. Describe in detail how a reliable value for the expansivity of mercury may be found by a method independent of the expansion of the containing vessel. Give the necessary theory.

A silica bulb of negligible expansivity holds 340.0 g of mercury at 0°C when full. Some steel balls are introduced and the remaining space is occupied at 0°C by 255.0 g of mercury. On heating the bulb and its contents to 100°C , 4.800 g of mercury overflow. Find the linear expansivity of the steel.

[Assume that the expansivity of mercury is $180 \times 10^{-6} \text{ K}^{-1}$.] (L.)

9. Describe an accurate method for determining the linear expansivity of a solid in the form of a rod. The pendulum of a clock is made of brass whose linear expansivity is 1.9×10^{-5} per deg C. If the clock keeps correct time at 15°C , how many seconds per day will it lose at 20°C ? (O. & C.)

10. A steel wire 8 metres long and 4 mm in diameter is fixed to two rigid supports. Calculate the increase in tension when the temperature falls 10°C . [Linear expansivity of steel = $12 \times 10^{-6} \text{ K}^{-1}$, Young's modulus for steel = $2 \times 10^{11} \text{ N m}^{-2}$.] (O. & C.)

11. How can you show that the density of water does not fall steadily as the temperature is raised from 0°C to 100°C ? What does your experiment indicate about the expansion of water? What importance has this result in nature? (C.)

12. Why is mercury used as a thermometric fluid? Compare the advantages and disadvantages of the use of a mercury-in-glass thermometer and a platinum resistance thermometer to determine the temperature of a liquid at about 300°C .

A dilatometer having a glass bulb and a tube of uniform bore contains 150 g of mercury which extends into the tube at 0°C . How far will the meniscus rise up the tube when the temperature is raised to 100°C if the area of cross-section of the bore is 0.8 mm^2 at 0°C ? Assume that the density of mercury at 0°C is 13.6 g cm^{-3} , that the expansivity of mercury is $1.82 \times 10^{-4} \text{ K}^{-1}$, and that the linear expansivity of glass is $1.1 \times 10^{-5} \text{ K}^{-1}$. (N.)

13. Describe in detail how you would determine the linear expansivity of a metal rod or tube. Indicate the chief sources of error and discuss the accuracy you would expect to obtain.

A steel cylinder has an aluminium alloy piston and, at a temperature of 20°C when the internal diameter of the cylinder is exactly 10 cm , there is an all-round clearance of 0.05 mm between the piston and the cylinder wall. At what temperature will the fit be perfect? (The linear expansivity of steel and the aluminium alloy are 1.2×10^{-5} and $1.6 \times 10^{-5} \text{ K}^{-1}$ respectively.) (O. & C.)

14. Explain the statement: *the absolute expansivity of mercury is $1.81 \times 10^{-4} \text{ K}^{-1}$* . Describe an experiment to test the accuracy of this value. Why is a knowledge of it important?

Calculate the volume at 0°C required in a thermometer to give a degree of length 0.15 cm on the stem, the diameter of the bore being 0.24 mm. What would be the volume of this mercury at 100°C ? [The linear expansivity of the glass may be taken as $8.5 \times 10^{-6} \text{ K}^{-1}$.] (L.)

chapter twelve

Changes of State. Vapours

SOLID TO LIQUID: FUSION

The Solid State

Substances exist in the solid, liquid or gaseous state. In the *solid state*, a body has a regular, geometrical structure. Sometimes this structure gives the body a regular outward form, as in a crystal of alum; sometimes, as in a strand of wool, it does not. But X-rays can reveal to us the arrangement of the individual atoms or molecules in a solid; and whether the solid is wool or alum, we find that its atoms or molecules are arranged in a regular pattern. This pattern we call a *space-lattice*; its form may be simple, as in metals, or complicated, as in wool, proteins, and other chemically complex substances.

We consider that the atoms or molecules of a solid are vibrating about their mean positions in its space-lattice. And we consider that the kinetic energy of their vibrations increases with the temperature of the solid: its increase is the heat energy supplied to cause the rise in temperature. When the temperature reaches the melting-point, the solid liquefies. Lindemann has suggested that, at the melting-point, the atoms or molecules vibrate so violently that they collide with one another. The attractive forces between them can then no longer hold them in their pattern, the space-lattice collapses, and the solid melts. The work necessary to overcome the forces between the atoms or molecules of the solid, that is, to break-up the space-lattice, is the latent heat of melting or fusion.

The Liquid State

In the liquid state, a body has no form, but a fixed volume. It adapts itself to the shape of its vessel, but does not expand to fill it. We consider that its molecules still dart about at random, as in the gaseous state, and we consider that their average kinetic energy rises with the liquid's temperature. But we think that they are now close enough together to attract one another—by forces of a more-or-less gravitational nature. Any molecule approaching the surface of the liquid experiences a resultant force opposing its escape (see p. 128, Surface Tension) Nevertheless, some molecules do escape, as is shown by the fact that the liquid evaporates: even in cold weather, a pool of water does not last for ever. The molecules which escape are the fastest, for they have the greatest kinetic energy, and therefore the greatest chance of overcoming the attraction of the others. Since the fastest escape, the slower, which remain, begin to predominate: the average kinetic energy of the molecules falls, and the liquid cools. The faster a liquid evaporates, the colder it feels on the hand—petrol feels colder than water, water feels

colder than paraffin. To keep a liquid at constant temperature as it evaporates, heat must be supplied to it; the heat required is the latent heat of evaporation.

Melting and Freezing

When a solid changes to a liquid, we say it undergoes a *change of state* or *phase*. Pure crystalline solids melt and freeze sharply. If, for example, paradichlorobenzene is warmed in a test tube until it melts, and then allowed to cool, its temperature falls as shown in Fig. 12.1 (a).

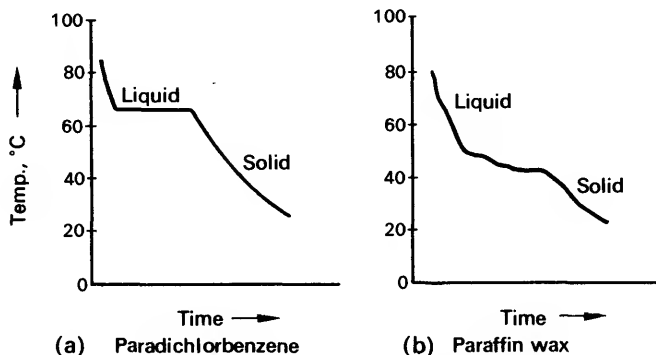


FIG. 12.1. Cooling curves, showing freezing.

A well-defined plateau in the cooling curve indicates the freezing (or melting) point. While the substance is freezing, it is evolving its latent heat of fusion, which compensates for the heat lost by cooling, and its temperature does not fall. An impure substance such as paraffin wax, on the other hand, has no definite plateau on its cooling curve; it is a mixture of several waxes, which freeze out from the liquid at slightly different temperatures (Fig. 12.1 (b)).

Supercooling

If we try to find the melting-point of hypo from its cooling curve, we generally fail; the liquid goes on cooling down to room temperature. But if we drop a crystal of solid hypo into the liquid the temperature rises to the melting-point of hypo, and the hypo starts to freeze. While the hypo is freezing, its temperature stays constant at the melting-point; when all the hypo has frozen, its temperature starts to fall again (Fig. 12.2).

The cooling of a liquid below its freezing-point is called *supercooling*; the

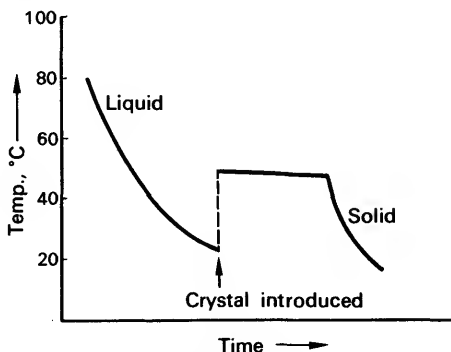


FIG. 12.2. Cooling curve of hypo.

molecules of the liquid lose their kinetic energy as it cools, but do not take up the rigid geometric pattern of the solid. Shaking or stirring the liquid, or dropping grit or dust into it, may cause it to solidify; but dropping in a crystal of its own solid is more likely to make it solidify. As soon as the substance begins to solidify, it returns to its melting (or freezing) point. *The melting-point is the only temperature at which solid and liquid can be in equilibrium.*

No one has succeeded in warming a solid above its melting-point—or, if he has, he has failed to report his success. We may therefore suppose that to superheat a solid is not possible; and we need not be surprised. For the melting-point of a solid is the temperature at which its atoms or molecules have enough kinetic energy to break up its crystal lattice: as soon as the molecules are moving fast enough, they burst from their pattern. On the other hand, when a liquid cools to its melting-point, there is no particular reason why its molecules should spontaneously arrange themselves. They may readily do so, however, around a crystal in which their characteristic pattern is already set up.

Pressure and Melting

The melting-point of a solid is affected by increase of pressure. If we run a copper wire over a block of ice, and hang a heavy weight from it, as in Fig. 12.3, we find that the wire slowly works through the block. It does not cut its way through, for the ice freezes up behind it; the pressure of the wire makes the ice under it melt, and above the wire, where the pressure is released, the ice freezes again. The freezing again after melting by pressure is called *regelation*.

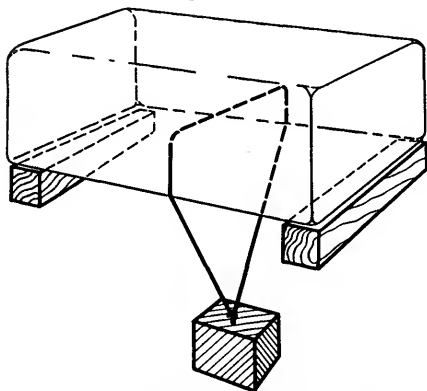


FIG. 12.3. Melting of ice under pressure.

This experiment shows that increasing the pressure on ice makes it melt more readily; that is to say, it *lowers the melting-point* of the ice. We can understand this effect when we remember that ice shrinks when it melts (see p. 209); pressure encourages shrinking, and therefore melting.

The fall in the melting-point of ice with increase in pressure is small: 0.0072°C per atmosphere. It is interesting, because it explains the slipperiness of ice; skates for example, are hollow ground, so that the pressure on the line of contact is very high, and gives rise to a lubricating film of water. Ice which is much colder than 0°C is not slippery, because to bring its melting-point down to its actual temperature would require a greater pressure than can be realized. Most substances swell on melting; an increase of pressure opposes the melting of such substances, and raises their melting-point.

Freezing of Solutions

Water containing a dissolved substance freezes below 0°C . The depression of the freezing-point increases at first with the concentration, but eventually reaches a maximum. The lowest freezing-point of common salt solution is -22°C , when the solution contains about one-quarter of its weight of salt. When a solution does freeze, pure ice separates out; an easy way of preparing pure water is therefore to freeze it, remove the ice, and then melt the ice. The water which is mixed with the ice in determining the ice-point of a thermometer must be pure, or its temperature will not be 0°C .

When ice and salt are mixed, the mixture cools below 0°C , but remains liquid. As the proportion of salt is increased, the temperature of the mixture falls, until it reaches a minimum at -22°C . A mixture of ice and salt provides a simple means of reaching temperatures below 0°C , and is called a 'freezing mixture'.

The phenomena of the freezing of solutions are important in chemistry, and particularly in metallurgy. We shall give a brief explanation of them later in this chapter.

LIQUID TO GAS: EVAPORATION

Evaporation differs from melting in that it takes place at all temperatures; as long as the weather is dry, a puddle will always clear up. In cold weather the puddle lasts longer than in warm, as the rate of evaporation falls rapidly with the temperature.

Solids as well as liquids evaporate. Tungsten evaporates from the filament of an electric lamp, and blackens its bulb; the blackening can be particularly well seen on the headlamp bulb of a bicycle dynamo set, if it has been frequently over-run through riding down-hill. The rate of evaporation of a solid is negligible at temperatures well below its melting-point, as we may see from the fact that bars of metal do not gradually disappear.

Saturated and Unsaturated Vapours

Fig. 12.4 (a) shows an apparatus with which we can study vapours and their pressures. A is a glass tube, about a metre long, dipping in a mercury trough and backed by a scale S. Its upper end carries a bulb B, which is fitted with three taps T_1 , of which T_1 and T_2 should be as close together as possible. Above T_1 is a funnel F. With T_1 closed but T_2 open, we evacuate the bulb and tube through T_3 , with a rotary pump. If the apparatus is clean, the mercury in A rises to the barometer height H . Meanwhile we put some ether in the funnel F. When the apparatus is evacuated, we close T_3 and T_2 . We now open and close T_1 , so that a little ether flows into the space C. Lastly, we open T_2 , so that the ether evaporates into the bulb B. As it does so, the mercury in A falls, showing that the ether-vapour is exerting a pressure (Fig. 12.4 (b)). If h is the new height of the mercury in A, then the pressure of the vapour in mm of mercury is equal to $H - h$.

By closing T_1 , opening and closing T_2 , and then opening T_1 again,

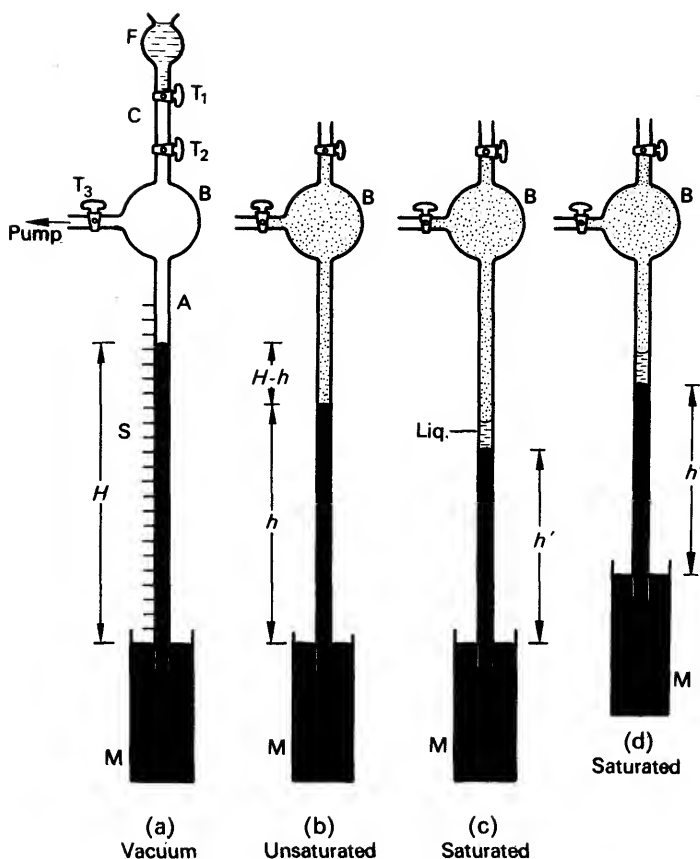


FIG. 12.4. Apparatus for studying vapours.

we can introduce more ether into the space B. At first, we find that, with each introduction, the pressure of the vapour, $H-h$, increases. But we reach a point at which the introduction of more ether does not increase the pressure, the height of the mercury column remains constant at h' . At this point we notice that liquid ether appears above the mercury in A (Fig. 12.4 (c)). We say that the vapour in B is now *saturated*; a *saturated vapour* is one that is in contact with its own liquid.

Before the liquid appeared in the above experiment, the pressure of the vapour could be increased by introducing more ether, and we say that the vapour in B was then *unsaturated*.

Behaviour of Saturated Vapour

To find out more about the saturated vapour, we may try to expand or compress it. We can try to compress it by raising the mercury reservoir M. But when we do, we find that the height h' does not change: the pressure of the vapour, $H-h'$, is therefore constant (Fig. 12.4 (d)). The only change we notice is an increase in the volume of liquid above

the mercury. We conclude, therefore, that reducing the volume of a saturated vapour does not increase its pressure, but merely makes some of it condense to liquid.

Similarly, if we lower the reservoir M, to increase the volume of the vapour, we do not decrease its pressure. Its pressure stays constant, but the volume of liquid above the mercury now decreases; liquid evaporates, and keeps the vapour saturated. If we increase the volume of the vapour until all the liquid has evaporated, then the pressure of the vapour begins to fall, because it becomes unsaturated (see Fig. 12.5 (a)).

Effect of Temperature: Validity of Gas-laws for Vapours

We cannot heat the apparatus of Fig. 12.4 through any great rise of temperature. But we can warm it with our hands, or by pointing an electric fire at it. If we do warm it, we find that the ether above the mercury evaporates further, and the pressure of the vapour increases. Experiments which we shall describe later show that the pressure of a saturated vapour rises, with the temperature, at a rate much greater than that given by Charles's law. Its rise is roughly exponential.

To Boyle's law, saturated vapours are indifferent: *their pressure is independent of their volume*. Unsaturated vapours obey Boyle's law roughly, as they also obey roughly Charles's law. Fig. 12.5 (a). Vapours, saturated and unsaturated, are gases in that they spread throughout their vessels; but we find it convenient to distinguish them by name from gases such as air, which obey Charles's and Boyle's laws closely. We shall elaborate this distinction later.

PROPERTIES OF SATURATED VAPOURS

Temperature °C	Water Pressure, mm mercury	Mercury	Ethyl Ether
		Pressure, mm mercury	Pressure, mm mercury
-20	0.784 (ice)		
-10	1.96 (ice)		
0	4.56	0.00016	112
10	9.20	0.00043	185
20	17.5	0.0011	291
30	31.7	0.0026	440
34.6			
40	55.1	0.0057	760
50	92.3	0.0122	921
60	149	0.0246	
70	234		1734
80	355	0.0885	
90	526		2974
100	760	0.276	
150	3569	2.88	4855
200	11647	17.8	
250	29770	75.8	
300	67620	249	
356.7		760	

Fig. 12.5 (b) shows the effect of heating a saturated vapour. More and more of the liquid evaporates, and the pressure rises very rapidly. As soon as all the liquid has evaporated, however, the vapour becomes unsaturated, and its pressure rises more sedately.

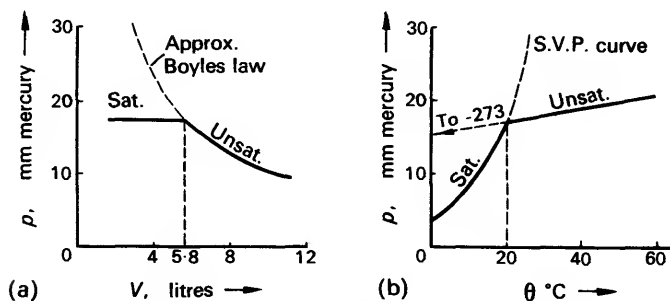


FIG. 12.5. Effect of volume and temperature on pressure of water vapour.

Fig. 12.6 (a) shows isothermals for a given mass of liquid and vapour at two temperatures, $\theta_1 = 10^\circ\text{C}$, and $\theta_2 = 21^\circ\text{C}$. The temperatures are chosen so that the saturated vapour pressure at θ_2 is double that at θ_1 . The absolute temperatures are $T_1 = 273 + \theta_1 = 283\text{ K}$, and $T_2 = 273 + \theta_2 = 294\text{ K}$. Because the saturated vapour pressure rises

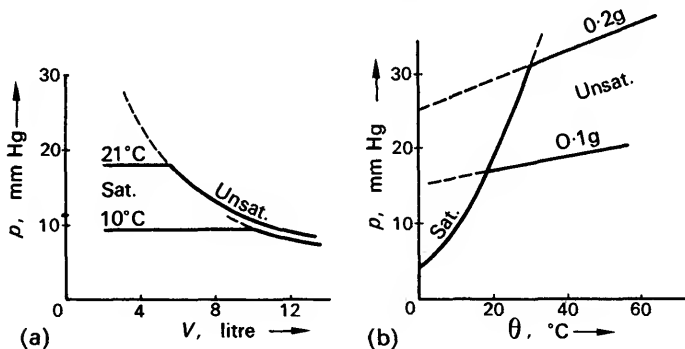


FIG. 12.6. Relationship between pressure, temperature, total mass, and volume, for water-vapour and liquid.

so rapidly with temperature, the absolute temperature T_2 is not nearly double the absolute temperature T_1 . Consequently the isothermals for the unsaturated vapour are fairly close together, as shown; and the transition from saturated to unsaturated vapour occurs at a smaller volume at the higher temperature.

Fig. 12.6 (b) shows pressure-temperature curves for a vapour, initially in contact with different amounts of liquid, in equal total volumes. The more liquid present, the greater is the density of the vapour when it becomes unsaturated, and therefore the higher the pressure and temperature at which it does so.

Kinetic Theory of Saturation

Let us consider a vapour in contact with its liquid, in an otherwise empty vessel which is closed by a piston (Fig. 12.7). The molecules of the vapour, we suppose, are rushing randomly about, like the molecules of a gas, with kinetic energies whose average value is determined by the temperature of the vapour. They bombard the walls of the vessel, giving rise to the pressure of the vapour, and they also bombard the surface of the liquid.

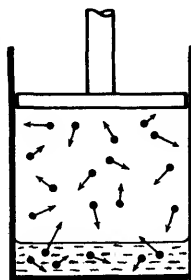


FIG. 12.7. Dynamic equilibrium.

The molecules of the liquid, we further suppose, are also rushing about with kinetic energies determined by the temperature of the liquid. The fastest of them escape from the surface of the liquid. At the surface, therefore, there are molecules leaving the liquid, and molecules arriving from the vapour. To complete our picture of the conditions at the surface, we suppose that the vapour molecules bombarding it are not reflected—as they are at the walls of the vessel—but are absorbed into the liquid. We may expect them to be, because we consider that molecules near the surface of a liquid are attracted towards the body of the liquid.

We shall assume that the liquid and vapour have the same temperature. Then the proportions of liquid and vapour will not change, if the temperature and the total volume are kept constant. Therefore, at the surface of the liquid, molecules must be arriving and departing at the same rate, and hence evaporation from the liquid is balanced by condensation from the vapour. This state of affairs is called a *dynamic equilibrium*. In terms of it, we can explain the behaviour of a saturated vapour.

The rate at which molecules leave unit area of the liquid depends simply on their average kinetic energy, and therefore on the temperature. The rate at which molecules strike unit area of the liquid, from the vapour, likewise depends on the temperature; but it also depends on the concentration of the molecules in the vapour, that is to say, on the density of the vapour. The density and temperature of the vapour also determine its pressure; the rate of bombardment therefore depends on the pressure of the vapour.

Now let us suppose that we decrease the volume of the vessel in Fig. 12.7 by pushing in the piston. Then we momentarily increase the density of the vapour, and hence the number of its molecules striking the liquid surface per second. The rate of condensation thus becomes greater than the rate of evaporation, and the liquid grows at the expense of the vapour. As the vapour condenses its density falls, and so does the rate of condensation. The dynamic equilibrium is restored when the rate of condensation, and the density of the vapour, have returned to their original values. The pressure of the vapour will then also have returned to its original value. *Thus the pressure of a saturated vapour is independent of its volume.* The proportion of liquid to vapour, however, increases as the volume decreases.

Let us now suppose that we warm the vessel in Fig. 12.7, but keep the piston fixed. Then we increase the rate of evaporation from the liquid, and increase the proportion of vapour in the mixture. Since the volume is constant, the pressure of the vapour rises, and increases the rate at which molecules bombard the liquid. Thus the dynamic equilibrium is restored, at a higher pressure of vapour. The increase of pressure with temperature is rapid, because the rate of evaporation of the liquid increases rapidly—almost exponentially—with the temperature. A small rise in temperature causes a large increase in the proportion of molecules in the liquid moving fast enough to escape from it.

Boiling

A liquid boils when its saturated vapour pressure is equal to the atmospheric pressure. To see that this is true, we take a closed J-shaped tube, with water trapped in its closed limb (Fig. 12.8 (a)). We heat the

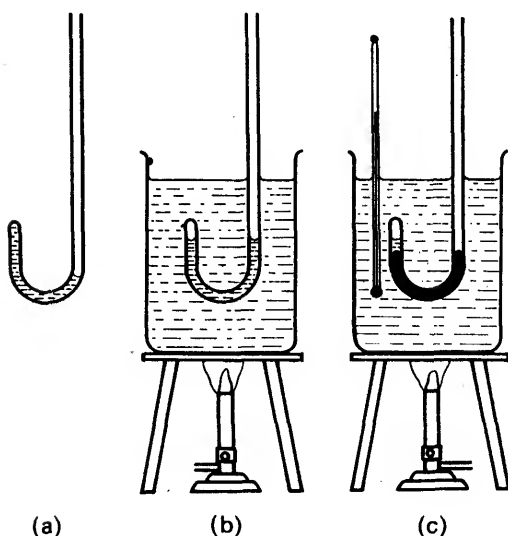


FIG. 12.8. Use of J-tube for boiling-point.

of which only a small quantity can be had. A few drops of the liquid are imprisoned by mercury in the closed limb of the tube, all entrapped air having been shaken out (Fig. 12.8 (c)). The tube is then heated in a bath, and the temperature observed at which the mercury comes to the same level in both limbs. The bath is warmed a little further, and then a second observation made as the bath cools; the mean of the two observations is taken as the boiling-point of the liquid.

Boiling differs from evaporation in that a liquid evaporates from its surface alone, but it boils throughout its volume. If we ignore the small hydrostatic pressure of the liquid itself, we may say that the pressure throughout a vessel of liquid is the atmospheric pressure. Therefore, when the saturated vapour pressure is equal to the atmospheric pressure, a bubble of vapour can form anywhere in the liquid.

Generally the bottom of the liquid is the hottest part of it, and bubbles form there and rise through the liquid to the surface. Just before the liquid boils, its bottom part may be at the boiling-point, and its upper part below. Bubbles of vapour then form at the bottom, rise in to the colder liquid, and then collapse. The collapsing gives rise to the singing of a kettle about to boil.

Further Consideration of Boiling

The account of boiling which we have just given is crude because in it we ignored the effect of surface-tension. Because of surface tension, a bubble can exist in a liquid only if there is an excess pressure inside it. If γ is the surface-tension of the liquid, and r the radius of the bubble, the excess pressure is $2\gamma/r$. If the bubble is formed at a depth h below the surface of the liquid, as in Fig. 12.9, the external pressure acting on it is

$$P = P_a + h\rho g,$$

where P_a = atmospheric pressure, ρ = density of liquid, g = acceleration of gravity.

Therefore a bubble, of radius r , can form at a depth h only if its vapour pressure, p , satisfies the equation

$$\begin{aligned} p &= P + \frac{2\gamma}{r} \\ &= P_a + h\rho g + \frac{2\gamma}{r}. \end{aligned}$$

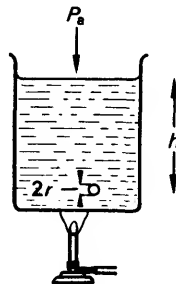


FIG. 12.9. Boiling.

If the radius r is small, the term $2\gamma/r$ is great, and the bubble cannot form unless the vapour pressure is considerably above atmospheric. In fact the equation shows that a bubble can never start from zero radius, because it would require an infinite vapour pressure to do so. Bubbles actually form on roughnesses in the vessel, or specks of solid suspended in the liquid. Very clean water in a very smooth beaker may not boil until it is well above 100°C ; its bubbles then grow violently, and the liquid 'bumps' in the beaker. A piece of broken pipe-clay prevents bumping, by presenting fine points for the bubbles to form on.

Thus the temperature of a boiling liquid is not definite—it depends on the conditions of boiling. But the temperature of the vapour is definite. The vapour escaping is in equilibrium with the liquid at the surface, and is at atmospheric pressure. Its temperature, therefore, is the temperature at which the saturated vapour pressure is equal to the atmospheric pressure. This idea is important in defining the upper fixed point of the temperature scale (p. 190). We say that the upper fixed point is the temperature of the steam from water boiling under a pressure of 760 mm mercury. We must not refer to the temperature of the water, and we must specify the atmospheric pressure because as we have just seen, it determines the temperature of the steam.

atmospheric, that is to say, when the temperature is above the normal boiling-point of the liquid, air is pumped into the reservoir R—with a bicycle pump—instead of drawing it out. The manometer M then shows the excess pressure, and

$$p = H + h.$$

With simple glass apparatus we cannot go far in this direction.

Boiling Point of a Solution

At a given pressure, the boiling-point of water containing a dissolved substance is higher than that of pure water. The temperature of the steam evolved from the solution, however, is the temperature of saturated steam at the prevailing pressure. Traces of dissolved substances in the water therefore do not affect the steam point in thermometry.

Since a liquid boils when its saturated vapour pressure is equal to the atmospheric pressure, we must conclude that dissolving a substance in water lowers its saturated vapour pressure, at a given temperature (Fig. 12.11). We may explain this by supposing that the molecules of the dissolved substance, which do not evaporate, hinder the escape of the molecules of the water.

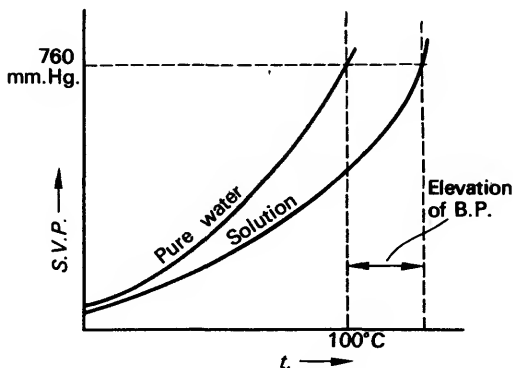


FIG. 12.11. Effect of solute on saturated vapour pressure.

The lowering of the vapour pressure of water by a dissolved solid gives striking support to the kinetic theory of evaporation. For measurements of the vapour pressure show that reduction does not depend on the nature of the solute; it depends only on the number of dissolved particles in the solution expressed as a fraction of the total number of particles (solute plus water molecules). In fact, if there are n solute particles to every $100 - n$ water molecules, then

$$\frac{\Delta p}{p} = \frac{n}{100},$$

where p is the saturated vapour pressure of water, and Δp is the lowering by solution. Thus the lowering simply depends on the number of particles hindering evaporation.

Effect of Altitude on Boiling Point

The pressure of the atmosphere decreases with increasing height above the earth's surface, because the thickness, and therefore the weight, of the belt of air above the observer decreases. The rate of fall in pressure is almost uniform over fairly small heights—about 85 mm mercury per km. But at great altitudes the rate of fall diminishes. At the height of Everest, 9000 m, the atmospheric pressure is about 280 mm of mercury. On account of the fall in atmospheric pressure, the boiling point of water falls with increasing height. Cooking-pots for use in high mountainous districts, such as the Andes, are therefore fitted with clamped lids. As the water boils, the steam accumulates in the pot, and its pressure rises above atmospheric. At about 760 mm mercury a safety valve opens, so that the pressure does not rise above that value, and the cooking is done at 100°C.

The fall in the boiling-point with atmospheric pressure gives a simple way of determining one's height above sea-level. One observes the steam point with a thermometer and hypsometer (p. 190). Knowing how the steam point falls with pressure, and how atmospheric pressure falls with increasing height, one can then find one's altitude. The hypsometer was, in fact, devised for this purpose, and takes its name from it; *hypsos* is Greek for height. Hypsometers have been carried up Himalayan peaks; and one was found by Scott and his companions in Amundsen's abandoned tent at the South Pole.

Variation of Latent Heat with Temperature

When we speak of the latent heat of evaporation of a liquid, we usually mean the heat required to vaporize unit mass of it at its normal boiling-point, that is to say, under normal atmospheric pressure. But since evaporation takes place at all temperatures, the latent heat has a value for every temperature. Regnault measured the latent heat of steam over a range of temperatures, by boiling water at controlled pressures, as in measuring its saturated vapour pressure. His apparatus was in principle similar to Berthelot's (Fig. 9.10); but he connected the outlet tube to an air reservoir, manometer, and pump, as in Fig. 12.10. Modern measurements give, approximately,

$$l = 2520 - 2.5\theta$$

where l is the specific latent heat in kJ kg^{-1} at $\theta^\circ\text{C}$.

Internal and External Latent Heats

The volume of 1 g of steam at 100°C is 1672 cm^3 . Therefore when 1 g of water turns into steam, it expands by 1671 cm^3 ; in doing so, it does work against the atmospheric pressure. The heat equivalent of this work is that part of the latent heat which must be supplied to the water to make it overcome atmospheric pressure as it evaporates; it is called the 'external latent heat'. The rest of the specific latent heat—the internal part—is the equivalent of the work done in separating the molecules, against their mutual attractions.

The work done, W , in the expansion of 1 g from water to steam is the product of the atmospheric pressure p and the increase in volume ΔV :

$$W = p \cdot \Delta V.$$

Normal atmospheric pressure corresponds to a barometer height H 760 mm. Hence, as on p. 228,

$$\begin{aligned} p &= g\rho H = 9.81 \times 13600 \times 0.76 \\ &= 1.013 \times 10^5 \text{ N m}^{-2} \end{aligned}$$

and $W = p \cdot \Delta V = 1.013 \times 10^5 \times 1671 \times 10^{-6}$ joule.

The external specific latent heat in joules is therefore

$$\begin{aligned} l_{ex} &= 1.013 \times 10^5 \times 1671 \times 10^{-6} \\ &= 170 \text{ J g}^{-1} = 170 \text{ kJ kg}^{-1}. \end{aligned}$$

This result shows that the external part of the specific latent heat is much less than the internal part. Since the total specific latent heat l is 2270 joule g^{-1} , the internal part is

$$\begin{aligned} l_{in} &= l - l_{ex} = 2270 - 170 \\ &= 2100 \text{ J g}^{-1} = 2100 \text{ kJ kg}^{-1}. \end{aligned}$$

Density of a Saturated Vapour

In any experiment to measure the density of a saturated vapour, the vessel containing the vapour must also contain some liquid, to ensure that the vapour is saturated. The problem is therefore to find how much of the total mass is vapour, and how much is liquid. Fig. 12.12 shows one method of solving this problem, due to Cailletet and Mathias.

A, B are two glass tubes, which have been calibrated with volume scales, and then evacuated. Known masses m_1 , m_2 of liquid are introduced into the tubes, which are then sealed off. The tubes are warmed to the same temperature in a bath, and the volumes of liquid V_{l1} , V_{l2} and of vapour V_{v1} , V_{v2} are observed. Then if ρ_l and ρ_v are the densities of liquid and vapour respectively:

$$m_1 = \rho_l V_{l1} + \rho_v V_{v1}$$

$$m_2 = \rho_l V_{l2} + \rho_v V_{v2}.$$

From these equations ρ_l can be eliminated, and ρ_v found. The equations can also, of course, be made to give ρ_l ; this method is useful for finding the density of a liquid 'gas'—e.g. liquid oxygen (p. 324).

Density of an Unsaturated Vapour

We have seen that the molecular weight of a gas, μ , is given very nearly by $\mu = 2\Delta$, where Δ is the density of the gas relative to that of hydrogen (p. 230).

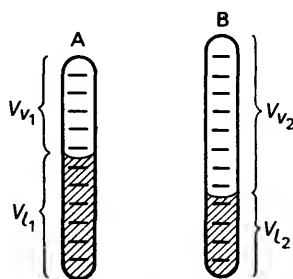


FIG. 12.12. Determination of density of a saturated vapour.

The proof that $\mu = 2\Delta$ depends on Avogadro's principle, which says that equal volumes of all gases at the same temperature and pressure contain equal numbers of molecules. This principle is true only of those gases which we normally call 'perfect'—which obey Boyle's and Charles's laws accurately (p. 225). It is not true of saturated vapours, but it is roughly true of vapours which are far from saturation. To find the molecular weight of a substance which is liquid at room temperature, therefore, we must vaporize it, and measure the density of its vapour when it is as far from saturation as we can conveniently get it.

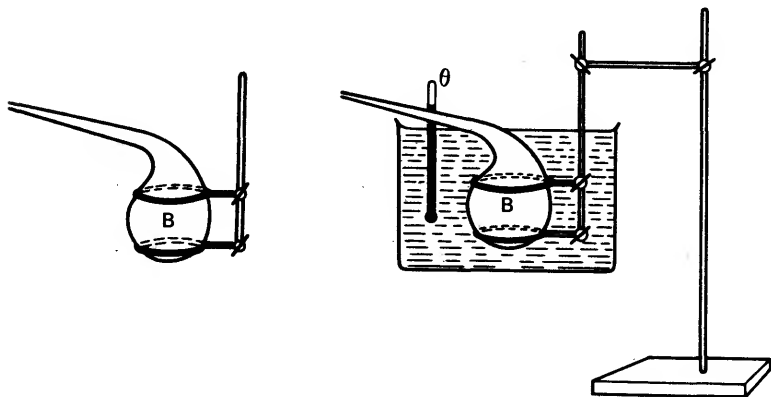


FIG. 12.13. Determination of density of an unsaturated vapour.

Several methods have been devised for doing this, one of which, Dumas's of 1827, is illustrated in Fig. 12.13. A glass bulb B, with a long thin stem, is weighed and then partly filled with liquid by warming and dipping. The amount of liquid introduced must be great enough to ensure that all the air in the bulb will be driven out by vapour when the liquid evaporates. The liquid is made to evaporate by plunging the bulb into a bath at a temperature θ about 40°C above its boiling-point. It then evaporates rapidly, and its vapour sweeps the air out of the bulb. When vapour has stopped coming out, the stem is sealed with a flame: the bulb now contains nothing but the vapour, at the temperature θ and under atmospheric pressure.

The bulb is removed, allowed to cool, dried, and weighed. The stem is then broken at the tip under water; since nearly all the vapour has condensed, at the room temperature, water rushes in and fills the bulb.

Let m_1 = mass of bulb full of air, at room temperature.

m_2 = mass of bulb full of vapour.

m_3 = mass of bulb full of water.

Since the mass of air in the bulb is negligible compared with that of water, we have, numerically,

$$\text{Volume of bulb in cm}^3, V_1, = m_3 - m_1 \text{ in g.}$$

The mass of air in the bulb at room temperature is

$$m_a = V\rho_a$$

where ρ_a is the density of air at room temperature and atmospheric pressure. The mass of the bulb itself is

$$m_b = m_1 - m_a$$

and the mass of vapour which filled it when hot is

$$m_v = m_2 - m_b$$

This mass of vapour occupied the volume V_1 at the temperature θ ; its density was therefore

$$\rho_v = \frac{m_v}{V_1}$$

Since the temperature was well above the boiling-point, the vapour was far from saturated; Boyle's and Charles's laws can therefore be used to reduce its density to s.t.p., for comparison with that of hydrogen.

WATER-VAPOUR IN THE ATMOSPHERE: HYGROMETRY

The water vapour in the atmosphere is important because it affects our comfort. Except in cold weather, we sweat continuously: the water in the sweat evaporates, draws its latent heat of evaporation from the skin, and so keeps us cool. Beads of sweat appear only when the water cannot evaporate as fast as it reaches the surface of the skin; we then feel uncomfortably hot.

On the other hand, if water evaporates from the skin too rapidly, the skin feels parched and hard; around the mucous membranes—at the mouth and nose—it tends to crack.

The rate at which water evaporates, from the skin or anywhere else, depends on the pressure of the water vapour surrounding it. If the water vapour above the skin is far from saturated, evaporation is swift. If the vapour is already saturated, water reaching the skin comes immediately into dynamic equilibrium with it; individual molecules are exchanged between liquid and vapour, but no mass of liquid is lost, and water accumulates.

The Partial Pressure of Atmospheric Water

The atmosphere contains other gases besides water-vapour, such as oxygen and nitrogen. In speaking of the water-vapour, therefore, we must refer to its 'partial pressure', as explained on p. 222.

Water-vapour in the atmosphere is also important because it affects the weather. Let us suppose that the atmosphere has a temperature of 20°C—a warm day—and that the water vapour in it has a partial pressure of 12 mm mercury. It will have a density of about 12 mg per litre. The density of *saturated* water vapour at 20°C is 17.3 mg per litre, and its pressure 17.5 mm mercury. The water vapour in the atmosphere is therefore not saturated.

Now let us suppose that the atmosphere cools to 14°C, without changing its composition. The 6°C fall in temperature will hardly affect the density of the water vapour, but it will bring the atmosphere to

saturation. For the pressure of saturated water vapour at 14°C is 12 mm mercury, and its density about 12 mg per litre. If the atmosphere cools any further, water vapour will condense out of it, forming drops of liquid water—that is, of fog or cloud.

Relative Humidity

The dampness of the atmosphere, besides affecting the weather and our comfort, is important also in storage and manufacture of many substances—tobacco and cotton, for example. From what we have said already, we can see that the important factor is not the actual proportion of water vapour in the atmosphere, but its nearness to saturation. In the above example, the density of the vapour remained almost constant, but we would have felt the atmosphere becoming much damper as it cooled from 20°C to 14°C.

The dampness of the atmosphere is expressed by its *relative humidity*, *R.H.*, which is defined as follows:

$$R.H. = \frac{\text{mass of water-vapour in a given volume of atmosphere}}{\text{mass of an equal volume of saturated water-vapour at the same temperature}} \quad (1)$$

In other words,

$$R.H. = \frac{\text{density of water-vapour in atmosphere}}{\text{density of saturated water-vapour at the same temperature}}$$

Because an unsaturated vapour roughly obeys Boyle's law, its density is roughly proportional to its pressure; the relative humidity as defined above is therefore roughly given by

$$R.H. = \frac{\text{partial pressure of water-vapour present}}{\text{S.V.P. at temperature of atmosphere}}$$

where S.V.P. stands for 'saturated vapour pressure'.

Before describing the methods of measurement, we must warn the reader against thinking that the atmosphere 'takes up' water vapour. The atmosphere is not a sponge. Water-vapour exists in it in its own right; and our knowledge of vapours makes us feel sure that, if we could live in an atmosphere of water-vapour alone, we would have just the same experiences of humidity as we now have in our happily richer surroundings.

Dew-point

In the evening, the earth cools more rapidly, by radiation, than the air above it. Then, on smooth surfaces such as metals, we often find a thin film of moisture. The surface has cooled to such a temperature that the water vapour in contact with it has become saturated, and has begun to condense. No fog has formed because the atmosphere in general is warmer than the cold solid, and the vapour in it is not saturated. The temperature of a cold surface on which dew just appears is called the *dew-point*: it is *the temperature at which the saturated vapour-pressure of water is equal to the partial pressure of the water-vapour present in the atmosphere*.

If we know the dew-point, we can find the corresponding pressure of saturated water-vapour, p_1 , from tables. From the same tables we can find the pressure of saturated water-vapour at the temperature of the atmosphere. If this is p_2 , then, from p. 310, the relative humidity, $R.H.$, is given by

$$R.H. = \frac{p_1}{p_2} \quad . \quad . \quad . \quad (2)$$

Dew-point Hygrometers

Hygrometry is the measurement of relative humidity, and a *hygrometer* is an instrument for measuring it. Many forms of dew-point hygrometer have been devised; we shall describe only the one due, like so much apparatus in Heat, to Regnault. It consists of two glass tubes, A, B, with silver-plated thimbles C, D cemented on to their lower ends. C contains ether, into which dips a thermometer T and a glass tube R (Fig. 12.14). To use the hygrometer, we first place a sheet of glass between it and ourselves, to prevent our breath from adding to the humidity of the air around it. We then gently blow air through the ether, by means of a scent-spray bulb connected through a rubber tube to R. (Alternatively, we may pass

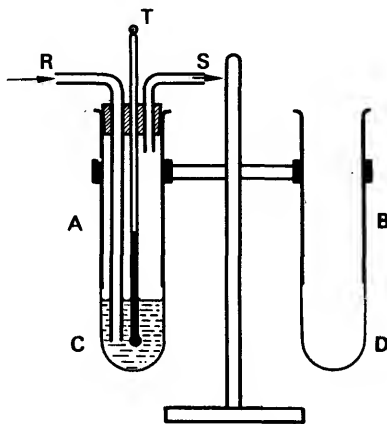


FIG. 12.14. Regnault's hygrometer.

in a gentle stream of coal-gas, and burn it at the end of a long tube connected to the outlet S.) The gas or air passing through the ether carries away its vapour, and makes it evaporate continuously. In doing so it cools, since it must provide its latent heat of evaporation from its own heat content. When the ether has cooled to the dew-point, a film of mist appears on the thimble C; the thimble D enables us to notice this more sharply, by contrast. At the moment when the dew appears, the thermometer T is read. The flow of air or gas through the ether is stopped, and the ether allowed to warm up. The temperature at which the dew vanishes is noted, and the mean of this and the temperature at which it appeared is taken as the dew-point, θ_1 say. By so doing we correct reasonably well for any difference of temperature between the ether and the outer surface of the tube. Lastly, we take the temperature of the room, θ_2 , often from a thermometer (not shown) in B, and look up the saturation pressures at θ_1 , θ_2 respectively. These are given in the following table. Then

$$\text{relative humidity} = \frac{\text{S.V.P. at } \theta_1}{\text{S.V.P. at } \theta_2} \times 100 \text{ per cent.}$$

PRESSURE AND DENSITY OF SATURATED WATER VAPOUR

$\theta^\circ\text{C}$	0	2	4	6	8	10	12	14	16	18	20
p , mm mercury	4.58	5.29	6.10	7.01	8.04	9.21	10.5	12.0	13.6	15.5	17.5
ρ , mg/litre (or g m^{-3})	4.84	5.54	6.33	7.22	8.21	9.33	10.6	12.0	13.5	15.2	17.1
$\theta^\circ\text{C}$	22	24	26	28	30	32	34	36	38	40	
p , mm mercury	19.8	22.3	25.1	28.3	31.7	35.5	39.8	44.4	49.5	55.1	
ρ , mg/litre (or g m^{-3})	19.2	21.5	24.1	26.9	30.0	33.5	37.2	41.3	45.8	51.1	

The Wet-and-dry-bulb Hygrometer

A piece of wet cloth feels cold, because the moisture evaporating from it takes latent heat, and cools the remaining liquid. This effect is used in the *wet-and-dry-bulb hygrometer*. A piece of muslin is tied round the bulb of a thermometer, and allowed to dip into a small

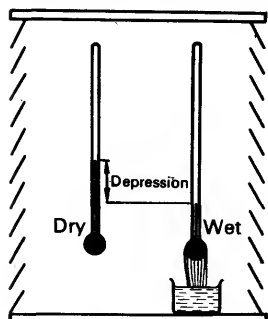


FIG. 12.15. Wet-and-dry bulb hygrometer.

jar of water. It is mounted, with a second, dry-bulb, thermometer, in a louvred draught-shield (Fig. 12.15). The rate at which water evaporates from the muslin increases as the relative humidity of the atmosphere falls; the cooling of the wet bulb therefore also increases. The greater the difference in reading of the two thermometers, the less is the relative humidity. By calibration against a chemical hygrometer, tables and charts have been prepared which give the relative humidity in terms of the thermometer readings.

WET-AND-DRY-BULB HYGROMETER TABLE
(Percentage relative humidity)

Dry-bulb reading $^\circ\text{C}$	Difference (depression of wet-bulb), $^\circ\text{C}$									
	1	2	3	4	5	6	8	10	12	14
0	82	65	48	31%						
5	85	72	58	45	32%					
10	88	76	65	54	44	34%				
15	90	80	71	61	52	44	27	12%		
20	91	83	74	66	59	51	37	24	12%	
25	92	84	77	70	63	57	44	33	22	12%
30		86	79	73	67	61	50	39	30	21%

The wet-and-dry-bulb hygrometer is not very reliable when used in a simple screen. It is more accurate if a steady stream of air is driven past it by a fan, or by whirling the thermometers around in a frame like a football-fan's rattle. The hygrometer is then said to be ventilated.

The Hair Hygrometer

Human hair expands in length in damp air. A hair hygrometer is one consisting of a bundle of hairs fixed to a spring at one end, and

wrapped round a spindle at the other. The expansion of the hair turns the spindle and moves a pointer over a scale, which is directly calibrated in relative humidities. Such instruments need to be recalibrated frequently, because the hair shows elastic fatigue.

THE BEHAVIOUR OF REAL GASES; CRITICAL PHENOMENA

A perfect, or ideal, gas is one which obeys Boyle's and Charles's law exactly, and whose internal energy is independent of its volume. No such gas exists, but at room temperature, and under moderate pressures, many gases approach the ideal closely enough for most purposes.

We shall consider now the departures of gases from perfection; in doing so we shall come to appreciate better the relationship between liquid, vapour, and gas, and we shall see how gases such as air can be liquefied.

Departures from Boyle's Law

In 1847 Regnault measured the volumes of various gases at pressures of several atmospheres. Using the apparatus of Fig. 12.16, he found that, to halve the volume of the gas, he did not have quite to double the pressure on it. The product pV , therefore, instead of being constant, decreased slightly with the pressure. He found one exception to this

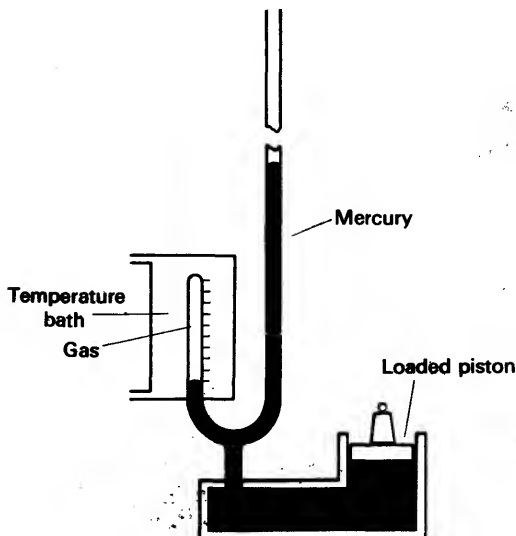


FIG. 12.16. Regnault's apparatus for isothermals at high pressure.

rule: hydrogen. By compressing the gases further, Regnault found the variation of pV with p at constant temperature, and obtained results which are represented by the early parts of the curves in Fig. 12.17.

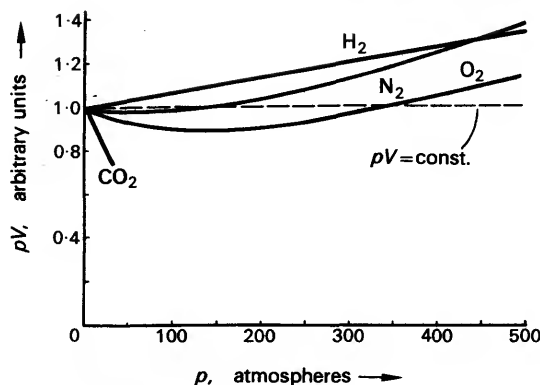


FIG. 12.17. Isothermals for various gases, at room temperature and high pressure.

The complete curves in the figure show some of the results obtained by Amagat in 1892. Amagat's apparatus for nitrogen is shown in Fig. 12.18. To get high pressures, he put the apparatus at the bottom of a coal-mine, and made the manometer tube out of rifle barrels, screwed together and standing up the shaft. He reached a pressure of 3000 atmospheres.

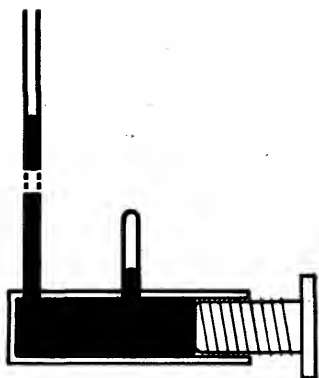


FIG. 12.18. Amagat's apparatus for isothermals at high pressure.

Having found the volume-pressure relationship for nitrogen, Amagat used it to measure high pressures in the laboratory, without having to resort to the mine. His method was similar to that of Andrews, which we are about to describe; by means of it he found the pressure-volume relationships for other gases.

Andrews' Work on Carbon Dioxide

In 1863 Andrews made experiments on carbon dioxide which have become classics. Fig. 12.19 shows his apparatus. In the glass tube A he trapped carbon dioxide above the pellet of mercury X. To do this, he started with the tube open at both ends and passed the gas through it for a long time. Then he sealed the end of the capillary. He introduced the mercury pellet by warming the tube, and allowing it to cool with the open end dipping into mercury. Similarly, he trapped nitrogen in the tube B.

Andrews then fitted the tubes into the copper casing C, which contained water. By turning the screws S, he forced water into the lower parts of the tubes A and B, and drove the mercury upwards. The wide parts of the tubes were under the same pressure inside and out, and so were under no stress. The capillary extensions were strong enough to

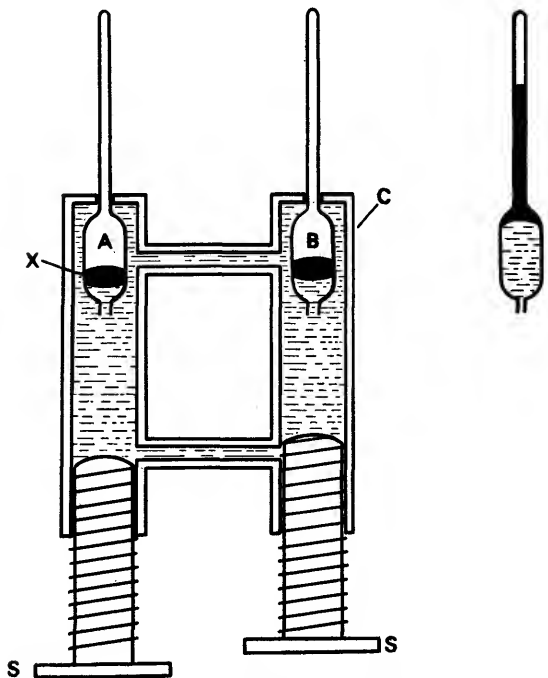


FIG. 12.19. Andrews's apparatus for isothermals of CO_2 at high pressures.

withstand hundreds of atmospheres. Andrews actually reached 108 atmospheres.

When the screws S were turned far into the casing, the gases were forced into the capillaries, as shown on the right of the figure, and greatly compressed. From the known volumes of the wide parts of the tubes, and the calibrations of the capillaries, Andrews determined the

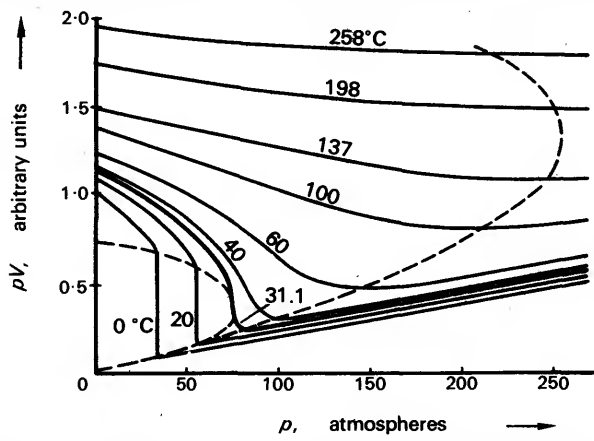


FIG. 12.20. Isotherms for CO_2 , as pV/p curves, at various temperatures. The small dotted loop passes through the ends of the vertical parts; the large dotted loop is the locus of the minima of pV .

volumes of the gases. He estimated the pressure from the compression of the nitrogen, assuming that it obeyed Boyle's law.

For work above and below room temperature, Andrews surrounded the capillary part of A with a water bath, which he maintained at a constant temperature between about 10°C and 50°C .

Fig. 12.20 shows some of Andrews's results, corrected for the departure of nitrogen from Boyle's law; it also shows the results of similar experiments over a wider range of temperature, by Amagat in 1893.

Critical Temperature

Before we can interpret Andrews's results for carbon dioxide, we must describe a simple experiment, made by Cagniard de la Tour in 1822.

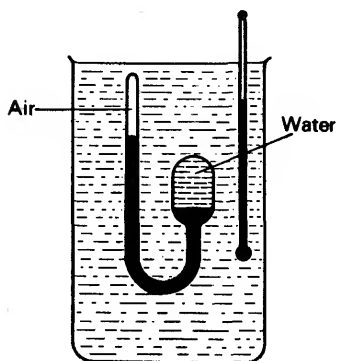


FIG. 12.21. Cagniard de la Tour's experiment.

De la Tour made a tube of strong glass, as shown in Fig. 12.21. In the bulb he had water, round the bend mercury, and at the top—where the tube was sealed off—air. He heated the tube in a bath to over 300°C . The expansion of the liquids was taken up by the compression of the air, from which de la Tour estimated the pressure; it went beyond 100 atmospheres. Above about 100°C he observed what we would expect; that a meniscus formed in bulb, showing that steam was present as well as water. But above about 300°C he noticed that the

meniscus vanished: that there was no observable distinction between liquid and vapour. The temperature at which the meniscus vanished he called the *critical temperature*.

If we consider the nature of a saturated vapour, the phenomenon of the critical temperature need not surprise us. For as its temperature rises a saturated vapour becomes denser, whereas a liquid becomes less dense. The critical temperature is, we may suppose, the temperature at which liquid and saturated vapour have the same density. Fig. 12.22 supports this view: it shows the results of measurements made on liquid oxygen by the method of Cailletet and Mathias (p. 307).

Behaviour of Carbon Dioxide near the Critical Point

Now let us turn to Andrews' isothermals for carbon dioxide. These are shown again, this time as a simple pressure-volume diagram, in Fig. 12.23. Let us consider the one for 21.5°C , ABCD. Andrews noticed that, when the pressure reached the value corresponding to B, a meniscus appeared above the mercury in the capillary containing the carbon dioxide. He concluded that the liquid had begun to form. From B to C, he found no change in pressure as the screws were turned, but simply a decrease in the volume of the carbon dioxide. At the same time the meniscus moved upwards, suggesting that the proportion of

liquid was increasing. At C the meniscus disappeared at the top of the tube, suggesting that the carbon dioxide had become wholly liquid. Beyond C the pressure rose very rapidly; this confirmed the idea that the carbon dioxide was wholly liquid, since liquids are almost incompressible.

Thus the part CBA of the isothermal for 21.5°C is a curve of volume

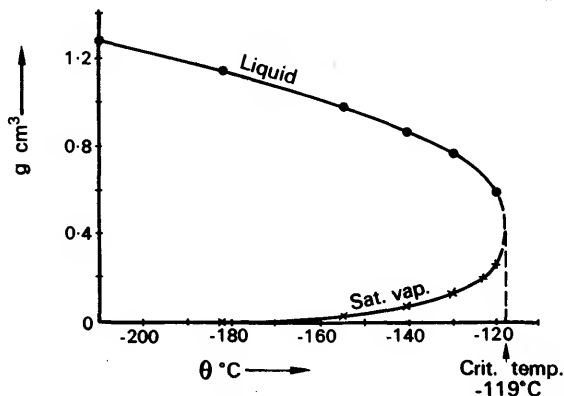


FIG. 12.22. Densities of liquid oxygen, and its saturated vapour.

against pressure for a liquid and vapour, showing saturation at B; it is like the isothermal for water given in Fig. 12.5 (a), p. 300. And the curve GFE is another such isothermal, for the lower temperature 13.1°C ; the two curves are like the two in Fig. 12.6 (a), p. 300.

The isothermal for 31.1°C has no extended plateau; it merely shows

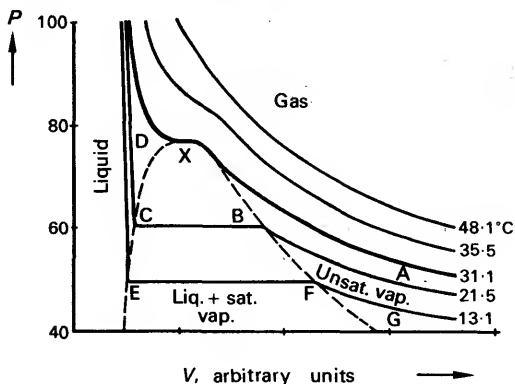


FIG. 12.23. Andrews's isothermals for CO_2 .

a point of inflection at X. At that temperature, Andrews observed no meniscus; he concluded that it was the critical temperature. The isothermals for temperature above 31.1°C never become horizontal, and show no breaks such as B or F. At temperatures above the critical, no transition from gas to liquid can be observed.

The isothermal for 48.1°C conforms fairly well to Boyle's law; even

when the gas is highly compressed its behaviour is not far from ideal.

The point X in Fig. 12.23 is called the critical point. The pressure and volume (of unit mass) corresponding to it are called the *critical pressure* and *volume*; the reciprocal of the critical volume is the critical density.

CRITICAL CONSTANTS OF GASES AND BOILING POINTS
(At atmospheric pressure)

Substance	Critical			Boiling-point °C
	Temperature °C	Pressure, atmospheres	Density kg m ⁻³	
Argon	-122	48	0.53×10^3	-186
Neon	-229	27	0.48	-246
Helium	-268	2.26	0.069	-269
Chlorine	146	76	0.57	-34
Hydrogen	-240	12.8	0.031	-253
Nitrogen	-146	33	0.31	-196
Oxygen	-118	50	0.43	-183
Air	-140	39	0.35	
Ammonia	130	115	0.24	-33.5
Carbon dioxide	31.1	73	0.46	-78.2
Ethylene	10	52	0.22	-102.7
Freon, CCl ₂ F ₂	112	40	0.56	
Sulphur dioxide	155	79	0.52	-10.8
Water	374	219	0.4	100

The above account of the phenomena near the critical point is over-simple, and may create the impression that these phenomena are fully understood. They are not; but this is not the place to say much about the matter. We may just point out that, even at temperature well above the critical, and when no meniscus can be seen, considerable differences of density can be found in a so-called gas. They have been shown by including, in a sealed tube of liquid and vapour, a number of small glass balls of different densities. When the tube was heated above the critical temperature, each ball floated at a point where the substance had a density equal to that of the ball.

Gases and Vapours

A gas above its critical temperature cannot be liquefied. Early attempts to liquefy gases such as air, by compression without cooling, failed; and the gases were wrongly called 'permanent' gases. We still, for convenience, refer to a gas as a vapour when it is below its critical temperature, and as a gas when it is above it. But the distinction is not the same as that between an ideal gas and one which is far from ideal. For a gas which is near its critical point, though it may be a little above its critical temperature, does not obey Boyle's law, as Fig. 12.23 shows. On the other hand, a vapour which is far from saturation obeys Boyle's law fairly well.

Refrigeration

The action of a refrigerator depends on the absorption of its latent heat by a liquid—the working substance—in evaporating. The working substance must be one whose vapour has a critical temperature above

normal atmospheric temperatures, so that it can be liquefied by compression alone. Common working substances are ammonia, carbon dioxide, sulphur dioxide, and specially developed compounds such as the two varieties of Freon: CCl_2F_2 , and $\text{C}_2\text{Cl}_2\text{F}_4$. The working substance is compressed by a pump, P, in Fig. 12.24, and passed through a metal pipe C; there the heat of compression is carried away by circulating water, and the substance liquefies. The liquid passes to a reservoir R. From the reservoir, liquid escapes through a throttle valve V into the coil D, which is connected to the low pressure side of the pump. The coil D lies round the walls of the space to be cooled (not shown).

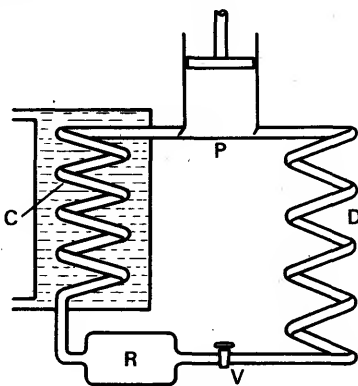


FIG. 12.24. Refrigeration.

When the liquid escapes from the reservoir, it starts to evaporate, because of the low pressure. It draws its latent heat from its own heat content, and cools. Not all of the liquid evaporates as it emerges, and the mixture of cool liquid and vapour passes round the metal coil D. If the atmosphere in the chamber containing D is warmer than the liquid, the liquid evaporates further. The latent heat which it requires is furnished by the surroundings of D, which are therefore cooled.

THE EQUILIBRIUM OF SOLID, LIQUID, VAPOUR

We have pointed out that solids as well as liquids evaporate (p. 297). A solid thus has saturated vapour over it, just as a liquid has, and the pressure of the saturated vapour depends on the temperature. The table on p. 299 shows the pressure of saturated water vapour over ice, at -10°C and -20°C .

The Triple Point

In Fig. 12.25, the curve AP relates the saturated vapour pressure of ice to its temperature; at any point on the curve, ice and water-vapour are in equilibrium. BP is the saturated-vapour-pressure curve of water; at any point on it, water and water-vapour are in equilibrium. CP is the curve relating the melting-point of ice with the pressure; at any point on it, ice and water are in equilibrium. The three curves meet at the point P, whose co-ordinates are $p = 4.6$ mm mercury, $\theta = 0.01^\circ\text{C}$. These are the only conditions in which ice, water, and water-vapour can exist together: if either the temperature or pressure is altered, at least one phase vanishes. If, when the pressure and temperature are altered, their new values happen to lie on one of the curves, then the two corresponding phases survive, liquid and solid along PC, for example. But if the new conditions lie in one of the three sectors of the diagram—say in PAC—then the only phase which survives is the one corresponding to that sector: solid, in PAC. The point P is called the *triple point*.

The curve AP, which gives the saturated vapour pressure of ice, is steeper at P than the curve BP for water. It is steeper because a solid evaporates less readily than a liquid—molecules escape from it less easily. Therefore the saturated-vapour-pressure of the solid falls more rapidly with the temperature.

Fig. 12.26 shows the triple point for carbon dioxide. Its co-ordinates are $p = 3800$ mm mercury, $\theta = -56.6^\circ\text{C}$. At atmospheric pressure, 760 mm mercury, therefore, solid carbon dioxide (CO_2) can be in equilibrium with its vapour, but not with liquid CO_2 . It is therefore dry, and in America it is called 'dry ice'; in England it is called 'carbon-dioxide snow'. At atmospheric pressure its temperature is -78.5°C , and it is much used as a coolant—in ice-cream trucks, for example.

Solid carbon dioxide is prepared by simply opening the valve of a cylinder containing carbon dioxide at high pressure. The gas rushes out, and does work

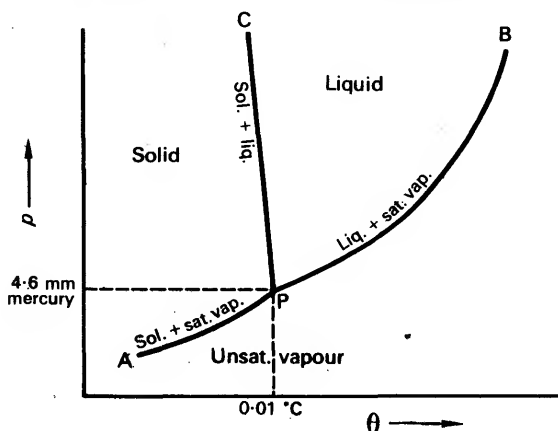


FIG. 12.25. Triple point for water (*not to scale*).

in acquiring kinetic energy of mass motion. Since the expansion is rapid, it is adiabatic, and the gas cools. As it does so, it goes over directly to the solid phase.

When solid carbon dioxide is warmed, it goes over directly into vapour. So, incidentally, do solid iodine and a few other substances, at atmospheric pressure. The change from solid to vapour is called sublimation. As the diagram shows, liquid CO_2 cannot exist at any temperature at all, if the pressure is below 3800 mm mercury (5.1 atmospheres).

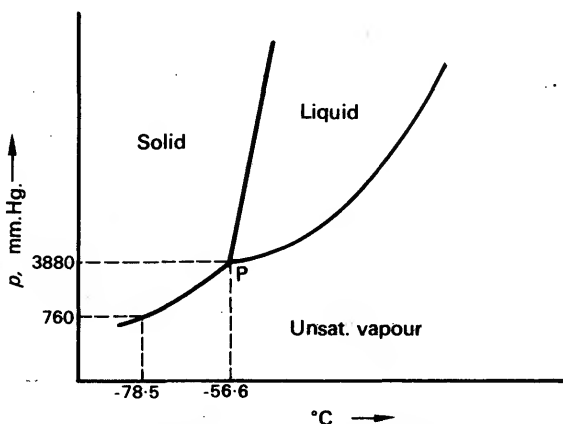


FIG. 12.26. Triple point for carbon dioxide (*not to scale*).

Freezing of Solutions

We have seen that dissolving a solid in water lowers its vapour pressure, and also its freezing-point (p. 297). To explain the lowering of the freezing-point, let us draw, as in Fig. 12.27, the curves of the saturated vapour pressures of ice,

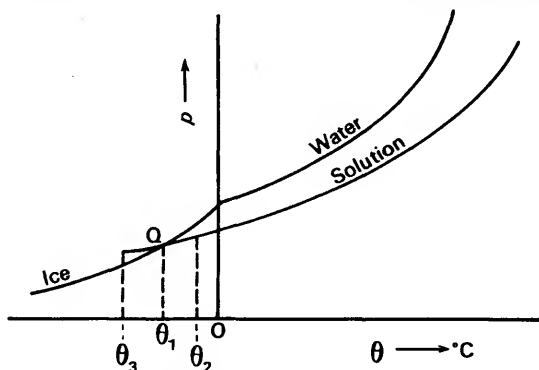


FIG. 12.27. Equilibrium between ice and a solution in water.

water and solution. We see that the curve for the solution cuts the ice curve at a point Q which corresponds to a temperature θ_1 , below 0°C . This is the only temperature at which ice and solution can be in equilibrium. At a higher temperature, θ_2 , ice has the higher vapour pressure: it therefore sublimates faster than water evaporates from the solution, and, on the whole, vapour from the ice condenses into the solution. At a lower temperature θ_3 , the solution has the higher vapour pressure; water therefore evaporates from it faster than the ice sublimates, and, on the whole, water from the solution condenses on the ice. Thus the temperature θ_1 is the freezing-point of the solution. It is the temperature at which solution and ice exchange water molecules one for one, and neither grows at the expense of the other.

We can now see why ice and salt, for example, form a freezing mixture. When salt is mixed with ice, it dissolves in the water clinging to the ice, and forms a solution. Since this is above its melting-point, being at 0°C , it has a lower saturation vapour pressure than the ice (Fig. 12.27). Therefore the ice sublimates and condenses in the solution. In effect, the ice becomes water. And in doing so it abstracts its latent heat of fusion from its surroundings. Thus the mixture changes from solid ice and salt to a liquid solution of salt, and its temperature falls below 0°C .

LIQUEFACTION OF GASES

If one of the so-called permanent gases, hydrogen or nitrogen, is to be liquefied, it must first be cooled below its critical temperature. There are three principal ways of doing this: (i) the gas may be passed through a cold bath containing a more easily liquefied gas, which is boiling at a reduced pressure and therefore has a very low temperature; (ii) the gas may be allowed to expand adiabatically and do work, losing its heat-energy in the process; (iii) the gas may be cooled by a method depending on the fact that, for a real gas, the internal energy is not independent of the volume.

The third of these is the commonest nowadays, and the only one

which we shall describe. First we must explain the phenomenon on which it depends.

The Joule-Kelvin Effect

We have already described, on p. 241, Joule's crude experiments on the expansion of a gas into a vacuum—a 'free expansion', as it is called. These experiments suggested that in such an expansion the

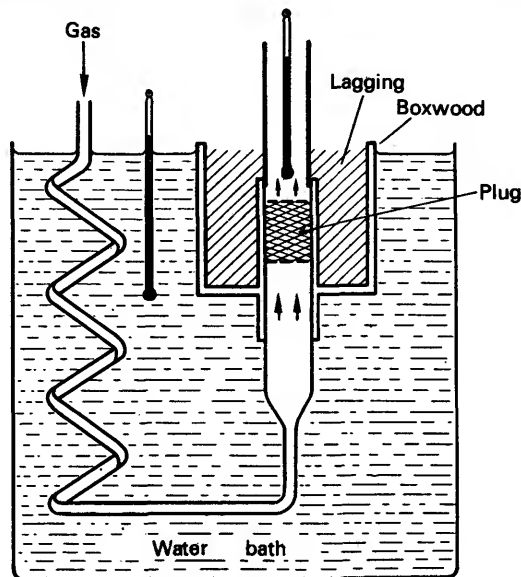


FIG. 12.28. Joule-Kelvin apparatus.

gas lost no internal energy, and therefore did no work. We concluded that the molecules of a gas had negligible attraction for one another, since otherwise work would have had to be done against their attractions whenever the gas expanded.

In 1852, Joule and Kelvin made more delicate experiments of essentially the same kind. They allowed a gas at high pressure to expand into a vacuum through a plug of cotton wool (Fig. 12.28). The plug prevented eddies from forming in the gas, so that the gas did not acquire any kinetic energy of motion in bulk. Neither did the gas do any external work, since it pushed back no piston. Nevertheless, Joule and Kelvin found that the gas was cooled slightly in its passage through the porous plug. Therefore work must have been done in separating its molecules; and this work must have been done at the expense of their kinetic energy, the heat-energy of the gas.

The magnitude of the cooling in the Joule-Kelvin effect depends on the temperature at which the gas enters the plug; for air at room temperature it has the order of 0.1°C per atmosphere pressure difference.

It is not essential for the gas to expand into a vacuum. Whenever a gas expands from high pressure to low, its volume increases, and some work is done against its inter-molecular attractions. If heat cannot

enter the gas, the work is done at the expense of the gas's internal energy, and the gas cools. The cooling is analogous to that which takes place in an adiabatic expansion, but in a normal adiabatic expansion most of the work is done externally against a piston (compare p. 251). An ideal gas would cool in an adiabatic expansion with external work, but not in a free expansion—it would show no Joule-Kelvin effect.

The Linde Process

The cooling of a gas in a free expansion is small, but Linde devised an ingenious arrangement for making it cumulative, and so producing a great temperature fall. His apparatus is shown diagrammatically in Fig. 12.29. When air is to be liquefied, it must first be freed of carbon dioxide and water, which would solidify and choke the pipes; both are removed by solid caustic soda in a vessel not shown in the figure. The pure air is compressed to about 150 atmospheres by the pump P, and the heat of compression is removed in the water-cooled copper coil C. The air then passes down the copper coil D, which runs within another copper coil E. It emerges through the nozzle N whose opening can be adjusted from outside. The nozzle lies inside a Dewar vessel or thermos flask F. The air expands on emerging, and is cooled by the Joule-Kelvin effect. It then passes upwards through the outer coil E, and as it does so, cools the incoming gas. The incoming gas is thus cooled before making its expansion, and after its expansion becomes cooler still. On escaping through E it cools the following gas yet further. Thus the cooling of the escaping gas continuously helps the cooling of the arriving gas, and the cooling is said to be *regenerative*. Eventually the gas emerging from the nozzle cools below the critical temperature; and since the actual pressure, 150 atmospheres, is well

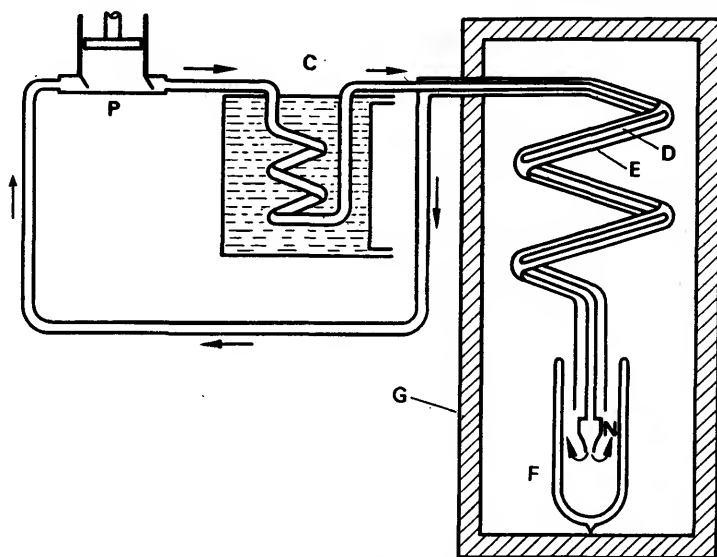


FIG. 12.29. Liquefaction of air.

above the critical pressure, 39 atmospheres, the gas liquefies and collects in the flask. The liquefier is heavily lagged with insulating material, G, to prevent heat coming in from the outside.

The reader should appreciate that the regenerative cooling takes place only in the double coil. At all stages in the process the air enters the inner coil at the temperature of the cooling water around C. But as time goes on it passes through ever-cooler gas coming up from the nozzle, until liquid begins to form and the system reaches equilibrium.

Liquid Nitrogen and Oxygen

As the table on p. 318 shows, the boiling-point of nitrogen, at atmospheric pressure, is -196°C , whereas that of oxygen is -183°C . When liquid air is exposed to the atmosphere, therefore, the nitrogen boils off faster and the proportion of liquid oxygen increases. The so-called liquid air sold commercially is mostly liquid oxygen. It is more dangerous than true liquid air, particularly if there is hydrogen about.

Hydrogen and Helium

At ordinary temperatures, the Joule-Kelvin effect is reversed for helium and hydrogen, that is, a free expansion causes warming. We cannot go into the explanation of that here, but we may say that it is connected with the fact that the pV/p curves of hydrogen and helium rise with increasing pressure, instead of falling at first (Fig. 12.17, p. 314).

These gases show a Joule-Kelvin cooling, however, if they are sufficiently cooled before the expansion. Hydrogen must be cooled below -83°C and helium below -240°C ; these temperatures are called the *inversion temperatures* of the gases.

Hydrogen can be cooled below its inversion temperature by passing it through a coil in liquid air before it enters the double coil of the liquefier. Helium must be passed through a coil in liquid hydrogen, boiling under reduced pressure.

EXAMPLES

1. Describe an experiment which demonstrates that the pressure of a vapour in equilibrium with its liquid depends on the temperature.

A narrow tube of uniform bore, closed at one end, has some air entrapped by a small quantity of water. If the pressure of the atmosphere is 760 mm of mercury, the equilibrium vapour pressure of water at 12°C and at 35°C is 10.5 mm of mercury and 42.0 mm of mercury respectively, and the length of the air column at 12°C is 10 cm, calculate its length at 35°C . (*L.*)

First part. See text.

Second part. For the given mass of air,

$$\frac{p_1 V_1}{T_1} = \frac{p_2 V_2}{T_2}$$

$$p_1 = 76 - 1.05 = 74.95 \text{ cm}, V_1 = 10, T = 273 + 12 = 285 \text{ K};$$

$$p_2 = 76 - 4.2 = 71.8 \text{ cm}, T_2 = 273 + 35 = 308 \text{ K};$$

$$\therefore \frac{74.95 \times 10}{285} = \frac{71.8 V_2}{308}$$

$$\therefore V_2 = \frac{74.95 \times 10 \times 308}{285 \times 71.8} = 11.3.$$

2. State Dalton's law of partial pressures; how is it explained on the kinetic theory? A closed vessel contains air, saturated water-vapour, and an excess of water. The total pressure in the vessel is 760 mm of mercury when the temperature is 25°C ; what will it be when the temperature has been raised to 100°C ? (Saturation vapour pressure of water at 25°C is 24 mm of mercury.) (C.)

First part. See text.

Second part. From Dalton's law, the pressure of the air at $25^\circ\text{C} = 760 - 24 = 736$ mm of mercury. Suppose the pressure is p mm at 100°C . Then, since pressure is proportional to absolute temperature for a fixed mass of air, we have

$$\frac{p}{736} = \frac{373}{298},$$

from which

$$p = 921 \text{ mm}.$$

Now the saturation vapour pressure of water at $100^\circ\text{C} = 760$ mm.

$$\therefore \text{total pressure in vessel} = 921 + 760 = 1681 \text{ mm mercury}.$$

3. Define *relative humidity* and *dew-point*. Describe an instrument with which the dew-point can be determined. The relative humidity in a closed room at 15°C is 60 per cent. If the temperature rises to 20°C , what will the relative humidity become? On what assumptions is your calculation based? (Saturation vapour pressure of water-vapour at $15^\circ\text{C} = 12.67$ mm of mercury, at $20^\circ\text{C} = 17.36$ mm.) (L.)

First part. See text.

Second part. Suppose p is the actual water vapour pressure in mm mercury in the air at 15°C . Then

$$\frac{p}{12.67} = \text{relative humidity} = 60 \text{ per cent}.$$

$$\therefore p = \frac{60}{100} \times 12.67 = 7.60 \text{ mm}.$$

Assuming the pressure of the water-vapour is proportional to its absolute temperature, the pressure p_1 at 20°C is given by

$$\frac{p_1}{7.60} = \frac{273 + 20}{273 + 15}$$

$$\therefore p = \frac{7.60 \times 293}{288} = 7.73 \text{ mm}.$$

$$\therefore \text{relative humidity at } 20^\circ\text{C} = \frac{7.73}{17.36} \times 100 \text{ per cent} = 45 \text{ per cent}.$$

4. What is meant by *saturation pressure of water vapour*, *dew-point*? Describe briefly the principles underlying two different methods for the determination of the relative humidity in the laboratory.

A barometer tube dips into a mercury reservoir and encloses a mixture of air and saturated water vapour above the mercury column in the tube, the height of the column being 70 cm above the level in the reservoir. If the atmospheric pressure and the saturation pressure of water vapour are respectively 76 cm and 1 cm of mercury, determine the height of the column when the tube is depressed in the reservoir to reduce the air volume to half its initial value. (*L.*)

First part. The saturation pressure of water vapour is the pressure of water vapour in contact with water in a closed space. The dew-point is the temperature at which the air is just saturated with the water-vapour present. The different methods for measuring relative humidity concern the dew-point (Regnault) hygrometer and the wet-and-dry bulb hygrometer, discussed on pp. 311-2.

Second part. We apply the gas laws to the *air* only, as the mass of the air remains constant. From Dalton's law.

$$\begin{aligned}\text{pressure of air} &= \text{total pressure} - \text{pressure of water-vapour} \\ &= (76 - 70) - 1 = 5 \text{ cm mercury.}\end{aligned}$$

The volume of the air changes from V , say, to $V/2$. Hence the new pressure, p , of the air is given, from Boyle's law, by

$$\begin{aligned}5 \times V &= p \times \frac{V}{2} \\ \therefore p &= 10 \text{ cm.}\end{aligned}$$

$$\therefore \text{new total pressure of mixture of gases} = 10 + 1 = 11 \text{ cm.}$$

$$\therefore \text{new height of mercury column} = 76 - 11 = 65 \text{ cm.}$$

5. What is meant by the *relative humidity* of the air? Describe in detail a good method for finding it.

Air at 19.5°C has a relative humidity of 75 per cent. Calculate its dew-point and the mass of water vapour contained in 1 litre, being given that the boiling-points of water under pressure of 12 mm, 14 mm, 16 mm and 18 mm of mercury are 14°C , 16.45°C , 18.55°C and 20.45°C respectively. Assume that water vapour behaves as an ideal gas, that its density at s.t.p. is $0.00080 \text{ g per cm}^3$. (*N.*)

First part. The relative humidity of the air is the ratio of the mass of water-vapour in a given volume of the air to the mass of water-vapour required to saturate that volume. A good method for finding it is by the dew-point (Regnault) hygrometer, described on p. 311.

Second part.

$$\text{Relative humidity} = \frac{\text{s.v.p. of water at dew-point}}{\text{s.v.p. of water at } 19.5^\circ\text{C}} \times 100 \text{ per cent.}$$

$$\therefore 75 = \frac{\text{s.v.p. at dew-point}}{17 \text{ mm mercury}} \times 100$$

$$\therefore \text{s.v.p. at dew-point} = \frac{3}{4} \times 17 = 12.75 \text{ mm mercury.}$$

$$\begin{aligned}\therefore \text{dew-point} &= 14^\circ\text{C} + \frac{0.75}{2} \times 2.45^\circ\text{C} \\ &= 14.9^\circ\text{C,}\end{aligned}$$

as s.v.p. at 14°C = 12 mm, at 16.45°C = 14 mm.

To find the mass of water-vapour in 1 litre. The pressure of water-vapour =

s.v.p. at dew-point = 12.75 mm = 1.275 cm mercury; the absolute temperature = $273 + 19.5 = 292.5$ K.

$$\therefore \text{vol. in cm}^3 \text{ at s.t.p.} = 1,000 \times \frac{1.275}{76} \times \frac{273}{292.5}$$

$$\therefore \text{mass of water-vapour} = 1,000 \times \frac{1.275}{76} \times \frac{273}{292.5} \times 0.0008 \\ = 0.013 \text{ g.}$$

EXERCISES 12

1. Distinguish between a saturated and an unsaturated vapour. What is meant by *saturation vapour pressure*?

Describe an experiment to measure the saturation vapour pressure of water vapour for temperatures between 20°C and 100°C . Indicate graphically the results which would be obtained from such an experiment. (L.)

2. Explain concisely **four** of the following in terms of the simple kinetic theory of matter:

(a) energy must be supplied to a liquid to convert it to a vapour without change of temperature;

(b) when some water is introduced into an evacuated flask, some of the water at first evaporates, but subsequently, provided the temperature of the flask is kept constant, the volume of the water present remains unchanged;

(c) gases are generally poor conductors of heat compared with solids;

(d) when a gas, which is enclosed in a thermally insulated cylinder provided with a piston, is compressed by moving the piston, the temperature of the gas is raised;

(e) water can be heated by stirring. (O. & C.)

3. Distinguish between a *saturated* and an *unsaturated* vapour. Describe an experiment to investigate the effect of pressure on the boiling point of water and draw a sketch graph to show the general nature of the results to be expected.

A column of air was sealed into a horizontal uniform-bore capillary tube by a water index. When the atmospheric pressure was 762.5 torr (mm of Hg) and the temperature was 20°C , the air-column was 15.6 cm long; with the tube immersed in a water bath at 50°C , it was 19.1 cm long, the atmospheric pressure remaining the same. If the s.v.p. of water at 20°C is 17.5 torr, deduce its value at 50°C . (O. & C.)

4. What is meant by *saturation vapour pressure*? Describe an experiment to investigate the variation of the saturation pressure of water vapour with temperature.

Sketch the isothermal curve relating pressure and volume (a) for a mass of dry air at room temperature, (b) for water vapour at 100°C . (L.)

5. Explain the physical principles of a domestic refrigerator employing an evaporating liquid and give a labelled diagram showing its essential components. How may the temperature of the main storage compartment be regulated and what factors determine the lowest attainable temperature?

A certain refrigerator converts water at 0°C into ice at a maximum rate of 5 g per minute when the exterior temperature is 15°C . Assuming that the rate at which heat leaks into the refrigerator from its surroundings is proportional to the temperature difference between the exterior and interior and is $2.5 \text{ watt deg } ^\circ\text{C}^{-1}$, what is the maximum exterior temperature at which this refrigerator could just maintain a temperature of 0°C in the interior? [Specific latent heat of fusion of ice = 330 kJ kg^{-1} .] (O. & C.)

6. Use the simple kinetic theory of matter to answer the following questions:
- How do gases conduct heat?
 - Why does a liquid tend to cool when it evaporates?
 - Why does the boiling point of a liquid depend upon the external pressure?

Show that the pressure p , the density ρ , and the mean square molecular velocity $\overline{c^2}$ of an ideal gas are related by $p = \frac{1}{3}\rho\overline{c^2}$, stating any assumptions at the points where they become necessary in the proof. (*O. & C.*)

7. State Boyle's law and Dalton's law of partial pressures.

The space above the mercury in a Boyle's law apparatus contains air together with alcohol vapour and a little liquid alcohol. Describe how the saturation vapour pressure of alcohol at room temperature may be determined with this apparatus.

A mixture of air and saturated alcohol vapour in the presence of liquid alcohol exerts a pressure of 12.8 cm of mercury at 20°C. When the mixture is heated at constant volume to the boiling point of alcohol at standard pressure (i.e. 78°C), the vapour remaining saturated, the pressure becomes 86.0 cm of mercury. Find the saturation vapour pressure of alcohol at 20°C. (*L.*)

8. The saturation pressure of water vapour is 1.2 cm of mercury at 14°C and 2.4 cm of mercury at 25°C. Describe and explain the experiment you would perform to verify these data.

Explain qualitatively in terms of the kinetic theory of matter (a) what is meant by saturation vapour pressure, (b) the relative magnitudes of the saturation vapour pressures quoted.

Sketch a graph showing how the saturation pressure of water vapour varies between 0°C and 110°C. (*N.*)

9. Draw p against v curves for temperatures above, at, and below the critical temperature, for a real gas. Explain the significance of critical temperature.

Sketch graphs of pv against p at the same temperatures for a real gas. What would be the form of the pv against p curves for an ideal gas?

Describe briefly experiments which provide the data on which these curves for a real gas are based. (*O. & C.*)

10. Compare the properties of saturated and unsaturated vapours. By means of diagrams show how the pressure of (a) a gas, and (b) a vapour, vary with change (i) of volume at constant temperature, and (ii) of temperature at constant volume.

The saturation vapour pressure of ether vapour at 0°C is 185 mm of mercury and at 20°C it is 440 mm. The bulb of a constant volume gas thermometer contains dry air and sufficient ether for saturation. If the observed pressure in the bulb is 1000 mm at 20°C, what will it be at 0°C? (*L.*)

11. Explain the terms *relative humidity*, *dew-point*. Describe in detail an experiment to determine the dew-point.

Find the mass of air and water-vapour in a room of $20 \times 10 \times 5$ cubic metres capacity, the temperature being 20°C and the pressure 750 mm of mercury. Assume that the saturation pressure of water-vapour at the dew-point is 9.0 mm of mercury; that the density of dry air at s.t.p. is 1.30 kg m^{-3} ; that the density of water vapour is $\frac{8}{9}$ that of air under the same conditions of pressure and temperature. (*L.*)

12. Define *pressure of a saturated vapour*, *critical temperature*. Give an account of the isothermal curves for carbon dioxide at temperatures above and below its critical temperature.

Some liquid ether is sealed in a thick-walled glass tube, leaving a space containing only the vapour. Describe what is observed as the temperature of the tube and its contents is raised above 197°C, which is the critical temperature for

ether. (Assume that the vessel is strong enough to withstand the internal pressure.) (L.)

13. Define *dew-point* and explain what is meant by *relative humidity*. Describe how you would determine the dew-point of the atmosphere.

What is the relative humidity of an atmosphere whose temperature is 16.3°C if its dew-point is 12.5°C ? The following table gives the saturation pressure, p , of water-vapour in mm of mercury at various temperature, t .

$t^{\circ}\text{C}$	10	12	14	16	18	20
p mm	9.20	10.51	11.98	13.62	15.46	17.51

(N.)

14. Give briefly the principles of one method each for liquefying (a) chlorine, (b) hydrogen.

Describe a suitable container for liquid air. Point out carefully the physical principles involved. (C.)

15. Give an account of three methods used to obtain temperatures below 0°C . Describe a method for liquefying air. How must the procedure be modified in order to liquefy hydrogen? (N.)

16. Describe in detail the method you would use to find (a) the melting-point of lead, and (b) the boiling-point of brine.

An alloy of copper and silver is made with different percentage composition. For each mix the melting-point is measured, with the following results:—

M.P. deg C	960	810	760	740	760	790	900	1080
Per cent of copper in alloy	0	20	30	40	50	60	80	100

Plot a curve showing the relation between the melting-point in degrees Centigrade and the percentage of copper present. Comment on the curve. (L.)

chapter thirteen

Transfer of Heat

Conduction

If we put a poker into the fire, and hold on to it, then heat reaches us along the metal. We say the heat is *conducted*; and we soon find that some substances—metals—are good conductors, and others—such as wood or glass—are not. Good conductors feel cold to the touch on a cold day, because they rapidly conduct away the body's heat.

Temperature Distribution along a Conductor

In order to study conduction in more detail consider Fig. 13.1 (a), which shows a metal bar AB whose ends have been soldered into the

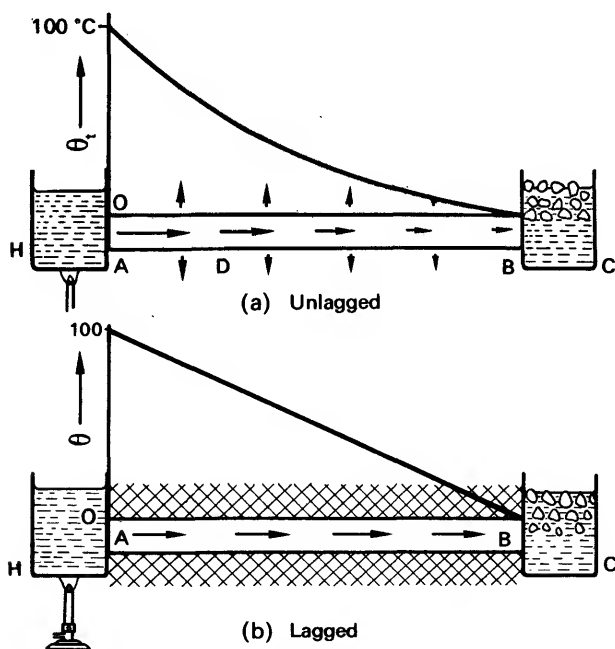


FIG. 13.1. Temperature fall along lagged and unlagged bars.

walls of two metal tanks, H, C; H contains boiling water, and C contains ice-water. Heat flows along the bar from A to B, and *when conditions are steady* the temperature θ of the bar is measured at points along its length; the measurements may be made with thermojunctions, not shown in the figure, which have been soldered to the rod. The curve in

the upper part of the figure shows how the temperature falls along the bar, less and less steeply from the hot end to the cold.

The figure 13.1 (b) shows how the temperature varies along the bar, if the bar is *well lagged* with a bad conductor, such as asbestos wool. It now falls uniformly from hot to cold.

The difference between the temperature distributions is due to the fact that, when the bar is unlagged, heat escapes from its sides, by convection in the surrounding air. Thus the heat flowing past D per second, is less than that entering the bar at A by the amount which escapes from the surface AD. The arrows in the figure represent the heat escaping per second from the surface of the bar, and the heat flowing per second along its length. The heat flowing per second along the length decreases from the hot end to the cold. But when the bar is lagged, the heat escaping from its sides is negligible, and the flow per second is constant along the length of the bar.

We thus see that the temperature gradient along a bar is greatest where the heat flow through it is greatest. We also see that the temperature gradient is uniform only when there is a negligible loss of heat from the sides of the bar.

Thermal Conductivity

Let us consider a very large thick bar, of which AB in Fig. 13.2 (i) is a part, and along which heat is flowing steadily. We suppose that the

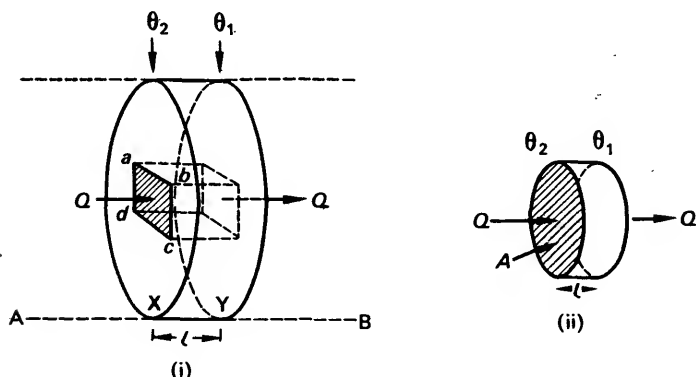


FIG. 13.2. Definition of thermal conductivity.

loss of heat from the sides of the bar is made negligible by lagging. XY is a slice of the bar, of thickness l , whose faces are at temperatures θ_2 and θ_1 . Then the *temperature gradient* over the slice is

$$\frac{\theta_2 - \theta_1}{l}.$$

We now consider an element $abcd$ of the slice of unit cross-sectional area, and we denote by Q the heat flowing through it per second. The value of Q depends on the temperature gradient, and, since some

substances are better conductors than others, it also depends on the material of the bar.

We therefore write

$$Q = k \frac{\theta_2 - \theta_1}{l}$$

where k is a factor depending on the material.

To a fair approximation the factor k is a constant for a given material; that is to say, it is independent of θ_2 , θ_1 , and l . It is called the *thermal conductivity* of the material concerned. To put its definition into words, we let $\theta_2 - \theta_1$ be 1°C , and l be 1 m, so that

$$Q = k.$$

We then say:

Consider a cube of material, whose faces are 1 m apart, and have a temperature difference of 1 deg C. If heat flows in the steady state through the cube at right angles to its faces, and none is lost from its sides, then the heat flow per unit area is numerically equal to the conductivity of the material.

This definition leads to a general equation for the flow of heat through any parallel-sided slab of the material, when no heat is lost from the sides of the slab. As in Fig. 13.2 (ii), we denote the cross-sectional area of the slab by A , its thickness by l , and the temperature of its faces by θ_1 and θ_2 . Then the heat Q flowing through it per second is

$$Q = \frac{kA(\theta_2 - \theta_1)}{l} \quad (1)$$

A useful form of this equation is

$$\frac{Q}{A} = k \frac{\theta_2 - \theta_1}{l} \quad (2)$$

or

heat flow per m^2 per second = conductivity \times temperature gradient. (2a)

In terms of the calculus, (2) may be re-written

$$\frac{1}{A} \frac{dQ}{dt} = -k \frac{d\theta}{dl} \quad (3)$$

the temperature gradient being negative since θ diminishes as l increases.

Units and Magnitude of Conductivity

Equation (2) enables us to find the unit of thermal conductivity. We have

$$k = \frac{Q/A (\text{J m}^{-2} \text{s}^{-1})}{(\theta_2 - \theta_1)/l (\text{K m}^{-1})}$$

Thus the unit of thermal conductivity = $\text{J s}^{-1} \text{m}^{-1} \text{K}^{-1}$, or since $\text{joule second}^{-1} = \text{watt (W)}$, the unit of k is $\text{W m}^{-1} \text{K}^{-1}$.

THERMAL CONDUCTIVITIES

SOLIDS

(Mean values, c. 0–100°C)

Substance	k $\text{W m}^{-1}\text{K}^{-1}$	Substance	k	Substance	k
Ag . . .	420	Asbestos . .	0.13	Ice . . .	2.1
Al . . .	210	Brick . . .	0.13	Marble . .	3.0
Cu . . .	380	Cardboard . .	0.21	(white)	
Fe pure . .	76	Cork . . .	0.42	Mica . . .	0.76
wrought . .	59	Cotton . . .	0.22	Paraffin wax .	0.25
Hg . . .	8	Cotton wool .	0.025	Silica (fused) .	1.4
Ni . . .	87	Ebonite . . .	0.17	Rubber (para)	0.19
Pb . . .	35	Felt . . .	0.038	Sand . . .	0.054
Pt . . .	71	Flannel . . .	0.097	Silk . . .	0.092
Brass . . .	109	Glass . . .	1.1	Slate . . .	2.0
Duralumin .	130	(window)		Wood . . .	c. 0.21
Steel . . .	46	Graphite . .	130		

LIQUIDS AND GASES

Liquid	k	Gas	k
Alcohol (25°C) . .	0.18	Air (0°C) . . .	0.024
Glycerine (20°C) . .	0.29	(100°C) . . .	0.032
Olive oil (0°C) . .	0.17	CO ₂ (0°C) . . .	0.015
Paraffin oil (0°C) .	0.13	H ₂ (0°C) . . .	0.17
Water (10°C) . . .	0.62	N ₂ (0°C) . . .	0.024
(80°C) . . .	0.67	O ₂ (0°C) . . .	0.024

To a rough approximation we may say that the conductivities of metals are about 1000 times as great as those of other solids, and of liquids; and they are about 10000 times as great as those of gases.

Effect of Thin Layer of Bad Conductor

Fig. 13.3 shows a lagged copper bar AB, whose ends are pressed against metal tanks at 0° and 100°C, but are separated from them by layers of dirt. The length of the bar is 10 cm or 0.1 m, and the dirt layers are 0.1 mm or 0.1×10^{-3} m thick. Assuming that the conductivity of dirt is 1/1000 that of copper, let us find the temperature of each end of the bar.

Suppose k = conductivity of copper,

A = cross-section of copper,

θ_2, θ_1 = temperature of hot and cold ends.

Since the bar is lagged, the heat flow per second Q is constant from end to end. Therefore,

$$Q = \frac{k}{1000} A \frac{100 - \theta_2}{0.1 \times 10^{-3}} = k A \frac{\theta_2 - \theta_1}{0.1} = \frac{k}{1000} A \frac{\theta_1 - 0}{0.1 \times 10^{-3}}.$$

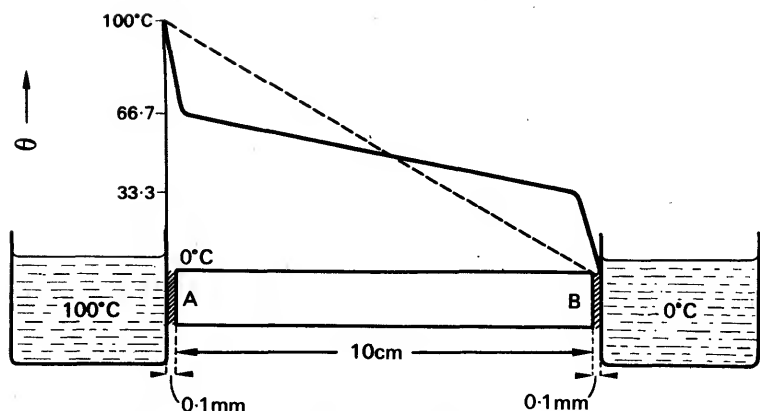


FIG. 13.3. Temperature gradients in good and bad conductors.

Dividing through k , these equations give

$$\frac{100 - \theta_2}{0.1} = \frac{\theta_2 - \theta_1}{0.1} = \frac{\theta_1}{0.1},$$

or

$$100 - \theta_2 = \theta_2 - \theta_1 = \theta_1,$$

whence

$$\theta_2 = 66.7^\circ\text{C},$$

$$\theta_1 = 33.3^\circ\text{C}.$$

Thus the total temperature drop, 100°C , is divided equally over the two thin layers of dirt and the long copper bar. The heavy lines in the figure show the temperature distribution; the broken line shows what it would be if there were no dirt.

Good and Bad Conductors

The foregoing example shows what a great effect a thin layer of a bad conductor may have on thermal conditions; 0.1 mm of dirt causes as great a temperature fall as 10 cm of copper. We can generalize this result with the help of equation (2a):

$$\text{heat flow/m}^2 \text{ s} = \text{conductivity} \times \text{temperature gradient}.$$

The equation shows that, if the heat flow is uniform, the temperature gradient is inversely proportional to the conductivity. If the conductivity of dirt is $1/1000$ that of copper, the temperature gradient in it is 1000 times that in copper; thus 1 mm of dirt sets up the same temperature fall as 1 m of copper. In general terms we express this result by saying that the dirt prevents a good thermal contact, or that it provides a bad one. The reader who has already studied electricity will see an obvious analogy here. The flow of heat can, in fact, be treated mathematically in the same way as the flow of electricity; we may say that a dirt layer has a high thermal resistance, and hence causes a great temperature drop.

Boiler plates are made of steel, not copper, although copper is about

eight times as good a conductor of heat. The material of the plates makes no noticeable difference to the heat flow from the furnace outside the boiler to the water inside it, because there is always a layer of gas between the flame and the boiler-plate. This layer may be very thin, but its conductivity is about $1/10000$ that of steel; if the plate is a centimetre thick, and the gas-film $1/1000$ centimetre, then the temperature drop across the film is ten times that across the plate. Thus the rate at which heat flows into the boiler is determined mainly by the gas.

If the water in the boiler deposits scale on the plates, the rate of heat flow is further reduced. For scale is a bad conductor, and, though it may not be as bad a conductor as gas, it can build up a much thicker layer. Scale must therefore be prevented from forming, if possible; and if not, it must from time to time be removed.

Badly conducting materials are often called *insulators*. The importance of building dwelling-places from insulating materials hardly needs to be pointed out. Window-glass is a ten-times better conductor than brick, and it is also much thinner; a room with large windows therefore requires more heating in winter than one whose walls are more modestly pierced. Wood is as bad a conductor (or as good an insulator) as brick, but it also is thinner. Wooden houses therefore have double walls, with an air-space between them; air is an excellent insulator, and the walls prevent convection. In polar climates, wooden huts must not be built with steel bolts going right through them; otherwise the inside ends of the bolts grow icicles from the moisture in the explorer's breath.

Measurement of High Conductivity: Metals

When the thermal conductivity of a metal is to be measured, two conditions must usually be satisfied: heat must flow through the specimen at a measurable rate, and the temperature gradient along the specimen must be measurably steep. These conditions determine the form of the apparatus used.

When the conductor is a metal, it is easy to get a fast enough heat flow; the problem is to build up a temperature gradient. It is solved by

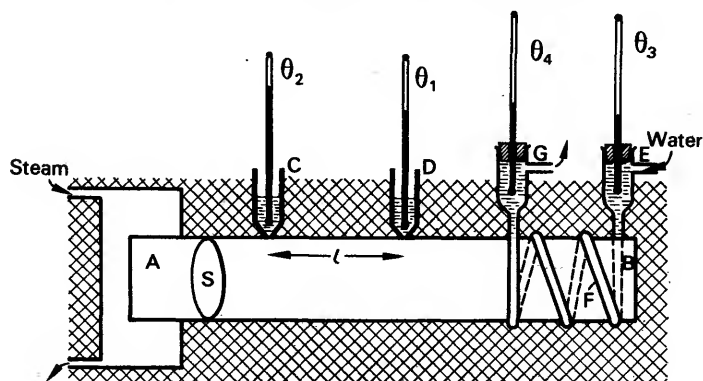


FIG. 13.4. Apparatus for thermal conductivity of a metal.

having as the specimen a bar long compared with its diameter. Fig. 13.4 shows the apparatus, which is due to Searle. AB is the specimen, about 4 cm diameter and 20 cm long. In one form of apparatus it is heated by steam at A, and cooled by circulating water at B. The whole apparatus is heavily lagged with felt. To measure the temperature gradient, thermometers are placed in the two mercury-filled cups C, D; the cups are made of copper, and are soldered to the specimen at a known distance apart. Alternatively, thermometers are placed in holes bored in the bar, which are filled with mercury. In this way errors due to bad thermal contact are avoided.

The cooling water flows in at E, round the copper coil F which is soldered to the specimen, and out at G. The water leaving at G is warmer than that coming in at E, so that the temperature falls continuously along the bar: if the water came in at G and out at E, it would tend to reverse the temperature gradient at the end of the bar, and might upset it as far back as D or C.

The whole apparatus is left running, with a steady flow of water, until all the temperatures have become constant: the temperature θ_2 and θ_1 , at C and D in the bar, and θ_4 and θ_3 of the water leaving and entering. The steady rate of flow of the cooling water is measured with a measuring cylinder and a stop-clock.

If A is the cross-sectional area of the bar and k its conductivity, then the heat flow per second through a section such as S is

$$Q = kA \frac{\theta_2 - \theta_1}{l}.$$

This heat is carried away by the cooling water; if a mass m of specific heat capacity c_w , flows through F in 1 second, the heat carried away is $mc_w(\theta_4 - \theta_3)$.

Therefore

$$kA \frac{\theta_2 - \theta_1}{l} = mc_w(\theta_4 - \theta_3).$$

With this apparatus we can show that the conductivity k is a constant over small ranges of temperature. To do so we increase the flow of cooling water, and thus lower the outflow temperature θ_4 . The gradient in the bar then steepens, and $(\theta_2 - \theta_1)$ increases. When the new steady state has been reached, the conductivity k is measured as before. Within the limits of experimental error, it is found to be unchanged.

Measurement of Low Conductivity: Non-metallic Solids

In measuring the conductivity of a bad conductor, the difficulty is to get an adequate heat flow. The specimen is therefore made in the form of a thin disc, D, about 10 cm in diameter and a few millimetres thick (Fig. 13.5 (a)). It is heated by a steam-chest C, whose bottom is thick enough to contain a hole for a thermometer.

The specimen rests on a thick brass slab B, also containing a thermometer. The whole apparatus is hung in mid air by three strings attached to B.

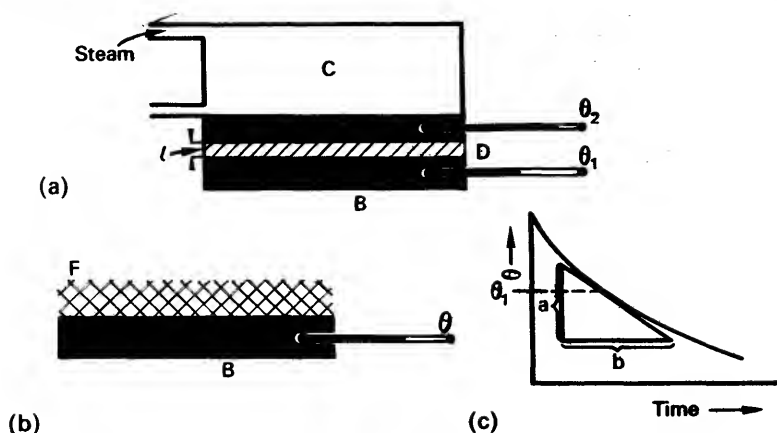


FIG. 13.5. Apparatus for thermal conductivity of a bad conductor.

To ensure good thermal contact, the adjoining faces of C, D and B must be flat and clean; those of C and B should be polished. A trace of vaseline smeared over each face improves the contact.

When the temperatures have become steady, the heat passing from C through D escapes from B by radiation and convection. Its rate of escape from B is roughly proportional to the excess temperature of B over the room (Newton's law). Thus B takes up a steady temperature θ_1 such that its rate of loss of heat to the outside is just equal to its gain through D. The rate of loss of heat from the sides of D is negligible, because their surface area is small.

This apparatus is derived from one due to Lees, and simplified for elementary work. If we use glass or ebonite for the specimen, the temperature θ_1 is generally about 70°C ; θ_2 is, of course, about 100°C . After these temperatures have become steady, and we have measured them, the problem is to find the rate of heat loss from B. To do this, we take away the specimen D and heat B directly from C until its temperature has risen by about 10°C . We then remove C, and cover the top part of B with a thick layer of felt F (Fig. 13.5 (b)). At intervals of a minute—or less—we measure the temperature of B, and afterwards plot it against the time (Fig. 13.5 (c)).

While the slab B is cooling it is losing heat by radiation and convection. It is doing so under the same conditions as in the first part of the experiment, because the felt prevents heat escaping from the top surface. Thus when the slab B passes through the temperature θ_1 , it is losing heat at the same rate as in the first part of the experiment. The heat which it loses is now drawn from its own heat content, whereas before it was supplied from C via D; that is why the temperature of B is now falling, whereas before it was steady. The rate at which B loses heat at the temperature θ_1 is given by:

$$\text{heat lost/second} = Mc \times \text{temperature fall/second},$$

where M , c are respectively the mass and specific heat capacity of the slab.

To find the rate of fall of temperature at θ_1 , we draw the tangent to the cooling curve at that point. If, as shown in Fig. 13.5 (c), its gradient at θ_1 would give a fall of a deg C in b seconds, then the rate of temperature fall is a/b deg C per second.

We then have, if A is the cross-sectional area of the specimen, l its thickness, and k its conductivity,

$$kA \frac{\theta_2 - \theta_1}{l} = Mc \frac{a}{b}.$$

Thus k can be calculated.

Liquids

In finding the conductivity of a liquid, the liquid must be heated at the top and cooled at the bottom, to prevent convection. Lees' apparatus is therefore suitable. The liquid is held in a narrow glass ring, R, Fig. 13.6, sandwiched between the plates (not shown) of the Lees' disc apparatus. Let k_g , k_l be the conductivities of the glass and liquid

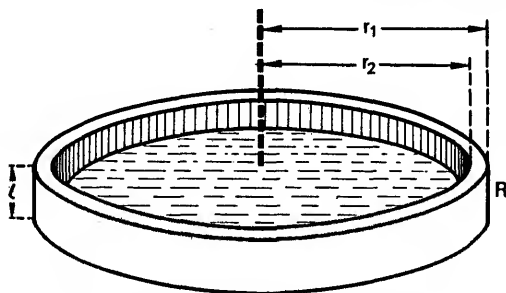


FIG. 13.6. Apparatus for finding thermal conductivity of a liquid.

respectively, and r_2 and r_1 the inner and outer radii of the ring. Then the downward heat flow per second is

$$Q = k_g \pi (r_1^2 - r_2^2) \frac{\theta_2 - \theta_1}{l} + k_l \pi r_2^2 \frac{\theta_2 - \theta_1}{l},$$

where θ_2 and θ_1 are the temperatures above and below the specimen and l is its thickness. The conductivity k_g need not be known; the heat flow through the ring may be determined in a preliminary experiment with the ring, but without the liquid.

Conduction through a Tube

The conductivity of glass tubing may be measured in the laboratory with the apparatus shown in Fig. 13.7 (a). The glass tube AB is surrounded by a steam jacket J, and water flows through it from A to B at a measured rate of m g/second. Thermometers measure the inflow and outflow temperatures of the water, θ_2 and θ_3 , which eventually become steady. In the steady state, the heat flowing through the walls of the tubing is equal to the heat carried away by the water, $mc_w(\theta_2 - \theta_3)$ joule/second.

To find the conductivity, we must know the area through which the heat flows. If r_1 , r_2 , are the inner and outer radii of the tube, and L is its length, then the areas of the inner and outer walls are $2\pi r_1 L$ and $2\pi r_2 L$ respectively (Fig. 13.7 (b)). If the tube is thin, we may take the area as constant and equal to its average value.

Thus
$$A = 2\pi L \frac{r_1 + r_2}{2}.$$

At the entrance of the tube, the temperature gradient is $(\theta_1 - \theta_2)/(r_2 - r_1)$, where θ_1 is the temperature of the steam; at the exit end the gradient is $(\theta_1 - \theta_3)/(r_2 - r_1)$.

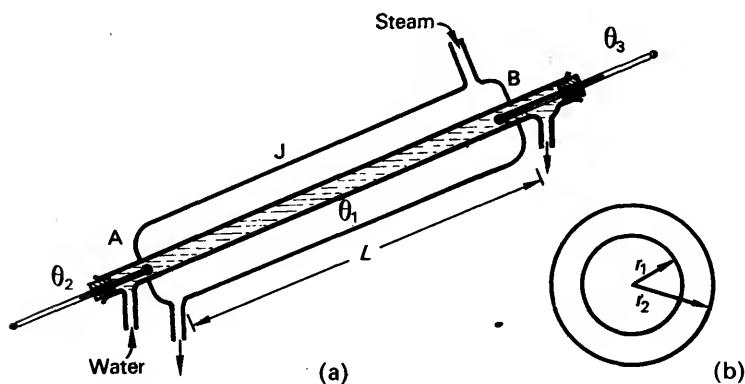


FIG. 13.7. Apparatus to measure conductivity of glass in form of tubing.

If θ_3 and θ_2 differ by not more than about 10°C , we may take the gradient as constant and equal to its average value:

$$\begin{aligned} \text{temperature-gradient} &= \frac{1}{2} \left(\frac{\theta_1 - \theta_2}{r_2 - r_1} + \frac{\theta_1 - \theta_3}{r_2 - r_1} \right) \\ &= \frac{\theta_1 - \frac{\theta_2 + \theta_3}{2}}{r_2 - r_1} \end{aligned}$$

The conductivity k therefore is given by

$$k \times 2\pi L \frac{r_1 + r_2}{2} \times \frac{\theta_1 - \frac{\theta_2 + \theta_3}{2}}{r_2 - r_1} = mc_w(\theta_3 - \theta_2) \quad (4)$$

The conductivity of *rubber tubing* can be found by a modification of this method. A measured length of the tubing is submerged in a calorimeter of water, and steam passed through for a measured time t . The rise in temperature of the water must be corrected for cooling, as in the measurement of the specific heat of a bad conductor (p. 203). The heat flow through the rubber is given by the left-hand side of equation (4) with θ_2 and θ_3 standing for the initial and final temperature of the

water. If m is the mass of water, and C the heat capacity of the calorimeter, the right-hand side of the equation is

$$(mc_w + C)(\theta_3 - \theta_2)/t.$$

Comparison of Conductivities

Fig. 13.8 (a) shows an apparatus due to Ingenhousz for comparing the conductivities of solids. Metal, wood, glass, and other rods, of equal lengths and cross-sections, are stuck into a tank through corks. The rods are painted with the same paint, and coated with wax over

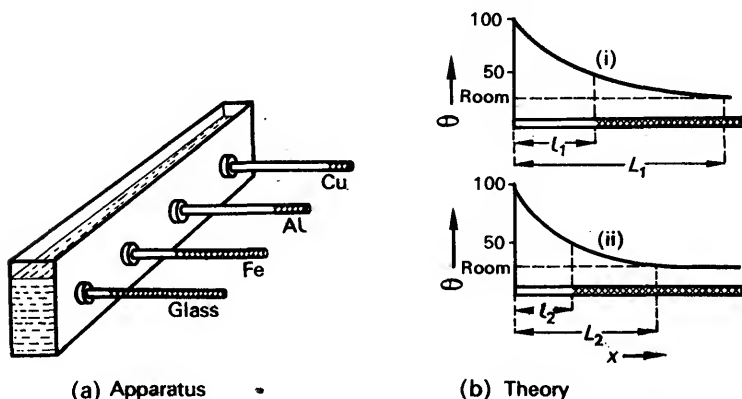


FIG. 13.8. Comparison of conductivities.

their whole projecting lengths. The tank is filled with water, leaks are stopped as well as possible, and the water is boiled. As the rods warm up, the wax melts off them. Eventually a steady state is reached, and the best conductor is the one from which wax has melted off the greatest length.

If the experiment is to give a quantitative comparison of the conductivities, the rods must be so long that the far end of each of them is at room temperature. Otherwise the following argument will not be true.

At the points where melted wax gives way to solid, each bar is at the melting-point of wax—let us call it 50°C . The temperature distributions along any two bars are therefore as shown in Fig. 13.8 (b); they have similar shapes, but at a given distance from the tank, (ii) is steeper than (i). If l_1 is the distance along (i) to the 50°C point, and l_2 the corresponding distance along (ii), then the curve (ii) is the same as curve (i) except that it is horizontally contracted in the ratio l_2/l_1 . Therefore the gradient of (ii), at any distance x , is steeper than that of (i) in the ratio l_1/l_2 . The temperature gradient at the tank end of each rod determines the rate at which heat flows into it from the hot water.

Now $Q \propto k \times \text{temp. grad. at hot end,}$

where k is the conductivity of the rod. Therefore, for rods (i) and (ii)

$$\begin{aligned} \frac{Q_1}{Q_2} &= \frac{k_1}{k_2} \times \frac{\text{temp. grad. at end of (i)}}{\text{temp. grad. at end of (ii)}} \\ &= \frac{k_1}{k_2} \times \frac{l_2}{l_1} \end{aligned} \quad (5)$$

The heat passing into a rod at the hot end escapes by convection from its sides. The lengths l_1 and l_2 respectively, of rods (i) and (ii), have the same average temperatures; 75°C . Over the lengths l_1 and l_2 , therefore, each rod loses heat at the same rate per unit area, since each has the same surface (p. 203). The heat lost from either rod per second, *between the 100°C and the 50°C points*, is therefore proportional to the area of the rod between those points. It is therefore proportional to the distance l between them.

Since the temperature curves differ only in scale, the distance to the 50°C point on either rod is proportional to the distance L to the point where the rod reaches room temperature (Fig. 13.8 (b)). Beyond this point, the bar loses no heat. Therefore, by the above argument, the distance L is proportional to the *total* heat lost per second by the bar. The distance L is therefore proportional to the total heat per second lost by the bar, and this heat is the heat Q entering the bar at the hot end.

Therefore $Q \propto L \propto l$

$$\text{or } \frac{Q_1}{Q_2} = \frac{L_1}{L_2} = \frac{l_1}{l_2}$$

$$\text{But we have seen that } \frac{Q_1}{Q_2} = \frac{k_1 l_2}{k_2 l_1} \quad (5)$$

$$\text{Therefore } \frac{k_1 l_2}{k_2 l_1} = \frac{l_1}{l_2}$$

$$\text{or } \frac{k_1}{k_2} = \frac{l_1^2}{l_2^2}$$

Thus the conductivity of a given rod is proportional to the square of the distance along it to the melting-point of the wax.

The Cracking of Glass

Glass is a bad conductor of heat. Therefore, when a piece of glass is heated in one place, the neighbouring parts of the glass do not at first warm up with it. Consequently they resist the expansion of the heated part, and the force set up cracks the glass (p. 262). To avoid cracking the glass, care must be taken to warm the whole region around the place to be made hot. Similarly, glass which has been heated must be made to cool slowly and uniformly by playing the flame over it now and then, for shorter and shorter times as it cools.

EXAMPLES

1. Calculate the quantity of heat conducted through 2 m^2 of a brick wall 12 cm thick in 1 hour if the temperature on one side is 8°C and on the other side is 28°C . (Thermal conductivity of brick = $0.13 \text{ W m}^{-1} \text{ K}^{-1}$.)

$$\text{Temperature gradient} = \frac{28-8}{12 \times 10^{-2}} ^\circ\text{C m}^{-1}.$$

Since 1 hour = 3600 seconds,

$$\begin{aligned} \therefore Q &= kAt \times \text{temperature gradient} \\ &= 0.13 \times 2 \times 3600 \times \frac{28-8}{12 \times 10^{-2}} \text{ joules} \\ &= 156\,000 \text{ J.} \end{aligned}$$

2. Define *thermal conductivity*. Describe and give the theory of a method of measuring the thermal conductivity of copper.

A sheet of rubber and a sheet of cardboard, each 2 mm thick, are pressed together and their outer faces are maintained respectively at 0°C and 25°C . If the thermal conductivities of rubber and cardboard are respectively 0.13 and $0.05 \text{ W m}^{-1} \text{ K}^{-1}$, find the quantity of heat which flows in 1 hour across a piece of the composite sheet of area 100 cm^2 . (L.)

First part. The thermal conductivity of a substance is the quantity of heat per second flowing in the steady state through opposite faces of a unit cube of the material when a temperature difference of 1 degree is maintained across these faces. The thermal conductivity of copper can be measured by Searle's method (p. 335).

Second part. We must first find the temperature, $\theta^{\circ}\text{C}$, of the junction of the rubber and cardboard. The temperature gradient across the rubber $= (\theta - 0)/2 \times 10^{-3}$; the temperature gradient across the cardboard $= (25 - \theta)/2 \times 10^{-3}$.

$$\therefore Q \text{ per second per m}^2 \text{ across rubber} = 0.13 \times (\theta - 0)/2 \times 10^{-3}$$

$$\text{and } Q \text{ per second per m}^2 \text{ across cardboard} = 0.05 \times (25 - \theta)/2 \times 10^{-3}.$$

But in the steady state the quantities of heat above are the same.

$$\therefore \frac{0.13(\theta - 0)}{2 \times 10^{-3}} = \frac{0.05(25 - \theta)}{2 \times 10^{-3}}$$

$$\therefore 13\theta = 125 - 5\theta$$

$$\therefore \theta = \frac{125}{18} = 7^{\circ}\text{C}.$$

$$\text{Now area} = 100 \text{ cm}^2 = 100 \times 10^{-4} \text{ m}^2.$$

$$\therefore Q \text{ through area in 1 hour (3600 seconds)}$$

$$= \frac{0.13 \times 100 \times 10^{-4} \times 7 \times 3600}{2 \times 10^{-3}}$$

$$= 16380 \text{ J}.$$

3. Define *thermal conductivity* and explain how you would measure its value for a poorly conducting solid.

In order to minimise heat losses from a glass container, the walls of the container are made of two sheets of glass, each 2 mm thick, placed 3 mm apart, the intervening space being filled with a poorly conducting solid. Calculate the ratio of the rate of conduction of heat per unit area through this composite wall to that which would have occurred had a single sheet of the same glass been used under the same internal and external temperature conditions. (Assume that the thermal conductivity of glass and the poorly conducting solid $= 0.63$ and $0.049 \text{ W m}^{-1} \text{ K}^{-1}$ respectively. (L.)

First part.

Second part. Let θ_1, θ_4 be the respective temperatures of the outer faces of the two glass sheets, and θ_2, θ_3 their respective junction temperatures with the solid between them. The thickness of glass $= 2 \times 10^{-3} \text{ m}$, that of the solid $= 0.3 \times 10^{-2} \text{ m}$. In the steady state, the quantity of heat per second per metre² is the same for each. Call this Q_1 . Then, for the first glass, since $Q_1 = k \times$ temperature gradient,

$$Q_1 = 0.63 \times \frac{\theta_1 - \theta_2}{2 \times 10^{-3}}$$

$$\therefore \theta_1 - \theta_2 = Q_1 \times \frac{2 \times 10^{-3}}{0.63} = \frac{2}{630} Q_1 \quad \quad \quad (i)$$

Similarly, for the solid,

$$\theta_2 - \theta_3 = Q_1 \times \frac{3 \times 10^{-3}}{0.049} = \frac{3}{49} Q_1 \quad (ii)$$

For the second glass,

$$\theta_3 - \theta_4 = Q_1 \times \frac{2 \times 10^{-3}}{0.63} = \frac{2}{630} Q_1 \quad (iii)$$

Adding (i), (ii), (iii) to eliminate θ_2 and θ_3 ,

$$\therefore \theta_1 - \theta_4 = Q_1 \left(\frac{4}{630} + \frac{3}{49} \right) \quad (1)$$

For a single sheet of glass and the same internal and external temperatures θ_1 and θ_4 respectively, the quantity of heat per second per metre², Q_2 say, is given, from (i), by

$$\theta_1 - \theta_4 = Q_2 \times \frac{2}{630} \quad (2)$$

Hence, from (1),
$$Q_1 \left(\frac{4}{630} + \frac{3}{49} \right) = Q_2 \times \frac{2}{630}$$

Simplifying,
$$\therefore \frac{Q_1}{Q_2} = \frac{14}{298} = 0.05 \text{ (approx.)}$$

RADIATION

Radiation

All heat comes to us, directly or indirectly, from the sun. The heat which comes directly travels through 150 million km of space, mostly empty, and travels in straight lines, as does the light: the shade of a tree coincides with its shadow. Both heat and light travel with the same speed because they are cut off at the same instant in an eclipse. Since light is propagated by waves of some kind we conclude that the heat from the sun is propagated by similar waves, and we say it is 'radiated'.

As we show later, radiation is more copious from a dull black body than from a transparent or polished one. Black bodies are also better absorbers of radiation than polished or transparent ones, which either allow radiation to pass through themselves, or reflect it away from themselves. If we hold a piece of white card, with a patch of black drawing ink on it, in front of the fire, the black patch soon comes to feel warmer than its white surround.

Reflection and Refraction

If, with either a convex lens or a concave mirror, we focus the sun's light on our skin, we feel heat at the focal spot. The heat from the sun has therefore been reflected or refracted in the same way as the light.

If we wish to show the reflection of heat unaccompanied by light, we may use two search-light mirrors, set up as in Fig. 13.9. At the focus of

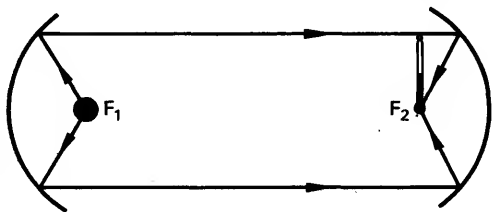


FIG. 13.9. Reflection of radiant heat.

one, F_1 , we put an iron ball heated to just below redness. At the focus of the other, F_2 , we put the bulb of a thermometer, which has been blackened with soot to make it a good absorber (p. 349). The mercury rises in the stem of the thermometer. If we move either the bulb or the ball away from the focus, the mercury falls back; the bulb has therefore been receiving heat from the ball, by reflection at the two mirrors. We can show that the foci of the mirrors are the same for heat as for light if we replace the ball and thermometer by a lamp and screen. (In practice we do this first, to save time in finding the foci for the main experiment.)

To show the refraction of heat apart from the refraction of light is more difficult. It was first done by the astronomer Herschel in 1800. Herschel passed a beam of sunlight through a prism, as shown diagrammatically in Fig. 13.10, and explored the spectrum with a sensitive

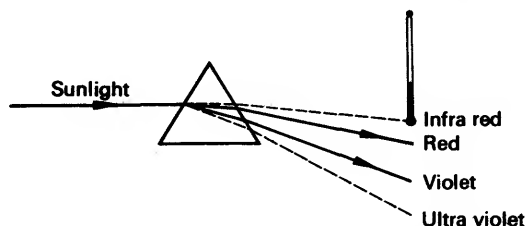


FIG. 13.10. Infra-red and ultra-violet (diagrammatic).

thermometer, whose bulb he had blackened. He found that in the visible part of the spectrum the mercury rose, showing that the light energy which it absorbed was converted into heat. But the mercury also rose when he carried the bulb into the darkened portion a little beyond the red of the visible spectrum; the sun's rays therefore carried energy which was not light.

Ultra-violet and Infra-red

The radiant energy which Herschel found beyond the red is now called *infra-red* radiation, because it is less refracted than the red. Radiant energy is also found beyond the violet and it is called *ultra-violet* radiation, because it is refracted more than the violet.

Ultra-violet radiation is absorbed by the human skin and causes sun-burn; more importantly, it stimulates the formation of vitamin D, which is necessary for the assimilation of calcium and the prevention of rickets. It is also absorbed by green plants; in them it enables water to combine with carbon dioxide to form carbohydrates. This process is called photo-synthesis; we have already, on p. 196, discussed its importance to animals and man. Ultra-violet radiation causes the emission of electrons from metals, as in photo-electric cells; and it excites a latent image on a photographic emulsion. It is harmful to the eyes.

Ultra-violet radiation is strongly absorbed by glass—spectacle-wearers do not sunburn round the eyes—but enough of it gets through

to affect a photographic film. It is transmitted with little absorption by quartz.

Infra-red radiation is transmitted by quartz, and rock-salt, but most of it is absorbed by glass. A little, that which lies near the visible red, passes fairly easily through glass—if it did not, Herschel would not have discovered it. When infra-red radiation falls on the skin, it gives the sensation of warmth. It is what we usually have in mind when we speak of heat radiation, and it is the main component of the radiation from a hot body; but it is in no essential way different from the other components, visible and ultra-violet radiation, as we shall now see.

Wavelengths of Radiation

In books on Optics, it is shown how the wavelength of light can be measured with a diffraction grating—a series of fine close lines ruled on glass. The wavelength ranges from 4000×10^{-10} m for the violet, to 7500×10^{-10} m for the red. The first accurate measurements of wavelength were published in 1868 by Angstrom, and in his honour a distance of 10^{-10} m is called an *Angstrom unit* (A.U.). The wavelengths of infra-red radiation can be measured with a grating made from fine wires stretched between two screws of close pitch. They range from 7500 A.U. to about 1 000 000 A.U. Often they are expressed in a longer unit than the Angstrom: this unit is the micron (μ m), which is $1/1000$ mm. Thus

$$1\mu\text{m} = 10^{-6} \text{ m} = 10^4 \text{ A.U.}$$

We denote wavelength by the symbol λ ; its value for visible light ranges from $0.4 \mu\text{m}$ to $0.75 \mu\text{m}$, and for infra-red radiation from $0.75 \mu\text{m}$ to about $100 \mu\text{m}$.

We now consider that X-rays and radio waves also have the same nature as light, and that so do the γ -rays from radio-active substances.

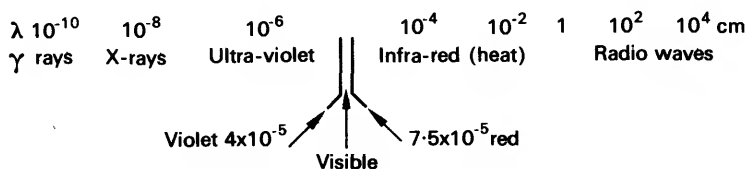


FIG. 13.11. The electromagnetic spectrum.

For reasons which we cannot here discuss, we consider all these waves to be due to oscillating electric and magnetic fields. Fig. 13.11 shows the range of their wavelengths: it is called a diagram of the *electromagnetic spectrum*.

Detection of Heat Radiation

A thermometer with a blackened bulb is a sluggish and insensitive detector of radiant heat. More satisfactory detectors, however, are

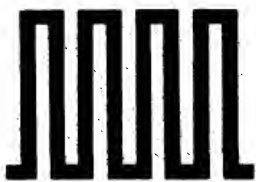


FIG. 13.12. Bolometer strip.

less direct; they are of two main kinds, both electrical. One kind consists of a long thin strip of blackened platinum foil, arranged in a compact zigzag (Fig. 13.12). On this the radiation falls. The foil is connected in a Wheatstone bridge, to measure its electrical resistance. When the strip is heated by the radiation, its resistance increases, and the increase is measured on the bridge. The instru-

ment was devised by Langley in 1881; it is called a bolometer, *bole* being Greek for a ray.

The other, commoner, type of radiation detector is called a thermopile (Nobili and Melloni, c. 1830). Its action depends on the electromotive force, which appears between the junctions of two different

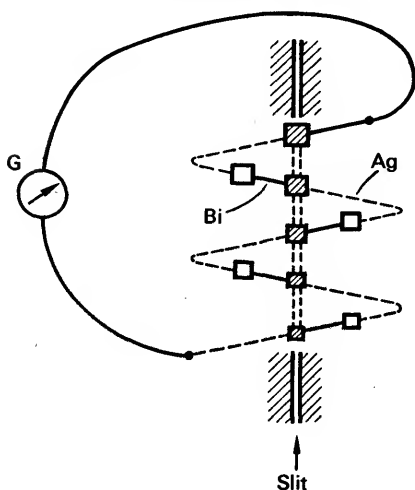


FIG. 13.13. Thermopile.

metals, when one junction is hot and the other cold. The modern thermopile is due to Coblentz (1913). It consists of many junctions between fine wires, as shown diagrammatically in Fig. 13.13; the wires are of silver and bismuth, 0.1 mm or less in diameter. Their junctions are attached to thin discs of tin, about 0.2 mm thick, and about 1 mm square. One set of discs is blackened and mounted behind a slit, through which radiation can fall on them; the junctions attached to them become the hot junctions of the thermopile. The other, cold, junctions are shielded from the radiation to

be measured; the discs attached to them help to keep them cool, by increasing their surface area.

Older types of thermopile are made from bars of metal about a millimetre thick. They are slow to warm up when radiation falls upon them, but are more rugged than the modern type.

When radiation falls on the blackened discs of a thermopile, it warms the junctions attached to them, and sets up an e.m.f. This e.m.f. can be measured with a potentiometer, or, for less accurate work, it can be used to deflect a galvanometer, G, connected directly to the ends of the thermopile (Fig. 13.13).

Reflection and Refraction observed with Thermopile: Inverse Square Law

With a thermopile and galvanometer, we can repeat Herschel's experiment more strikingly than with a thermometer. And with the

simple apparatus of Fig. 13.13 we can show that, when heat is reflected, the angle of reflection is equal to the angle of incidence. We can also show the first law of reflection; that the incident and reflected rays are in the same plane as the normal to the reflector at the point of incidence.

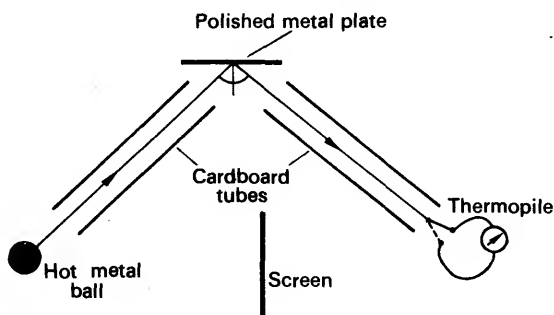


FIG. 13.14. Demonstration of reflection.

If heat is radiant energy, its intensity should fall off as the inverse square of the distance from a point source. We can check that it does so by setting up an electric lamp, with a compact filament, in a dark room preferably with black walls. When we put a thermopile at different distances from the lamp, the deflection of the galvanometer is found to be inversely proportional to the square of the distance.

If we wish to do this experiment with radiation that includes no visible light, we must modify it. Instead of the lamp, we use a large blackened tank of boiling water, A, and we fit the thermopile, B, with a conical mouthpiece, blackened on the inside. The blackening prevents any radiation from reaching the pile by reflection at the walls of the mouthpiece. We now find that the deflection of the galvanometer, G, does not vary with the distance of the pile from the tank, *provided that the tank occupies the whole field of view of the cone* (Fig. 13.15). The area S of the tank from which radiation can reach the thermopile is then proportional to the square of the distance d . And since the deflection is unchanged when the distance is altered, the total radiation from each element of S must therefore fall off as the inverse square of the distance d .

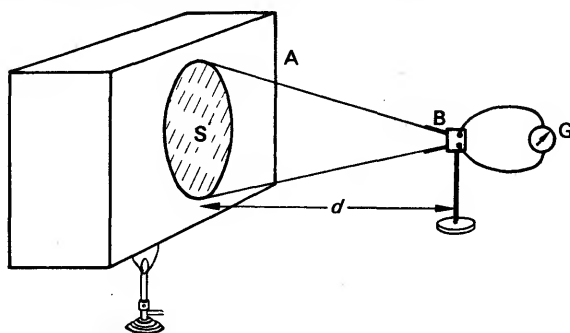


FIG. 13.15. Proof of inverse square law.

The Infra-red Spectrometer

Infra-red spectra are important in the study of molecular structure. They are observed with an infra-red spectrometer, whose principle is shown in Fig. 13.16. Since glass is opaque to the infra-red, the radiation

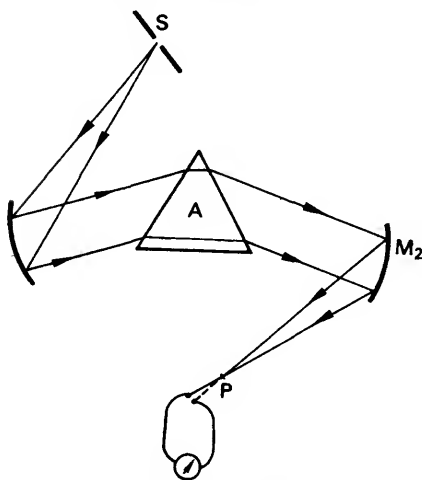


FIG. 13.16. Infra-red spectrometer.

is focused by concave mirrors instead of lenses; the mirrors are plated with copper or gold on their front surfaces. The source of light is a Nernst filament, a metal filament coated with alkaline-earth oxides, and heated electrically. The radiation from such a filament is rich in infra-red. A carbon arc, or a gas-mantle, may be used, however.

The slit S of the spectrometer is at the focus of one mirror which acts as a collimator. After passing through the rock-salt prism, A, the radiation is focused on to the thermopile P by the mirror M_2 , which re-

places the telescope of an optical spectrometer. Rotating the prism brings different wavelengths on to the slit; the position of the prism is calibrated in wavelengths with the help of a grating.

To a fair approximation, the deflection of the galvanometer is proportional to the radiant power carried in the narrow band of wavelengths which fall on the thermopile. If an absorbing body, such as a solution of an organic compound, is placed between the source and the slit, it weakens the radiation passing through the spectrometer, in the wavelengths which it absorbs. These wavelengths are therefore shown by a fall in the galvanometer deflection.

Reflection, Transmission, Absorption

Measurements whose description is outside our scope give the amount of radiant energy approaching the earth from the sun. At the upper limit of our atmosphere, it is about $8.0 \text{ J cm}^{-2} \text{ min}^{-1}$.

At the surface of the earth it is always less than this because of absorption in the atmosphere. Even on a cloudless day it is less, because the ozone in the upper atmosphere absorbs much of the ultra-violet.

In Fig. 13.17, XY represents a body on which radiant energy is falling. The symbol I denotes the latter's intensity: to fix our ideas we may take

$$\begin{aligned} I &= 4.0 \text{ joule per cm}^2 \text{ per minute} \\ &= \frac{4.0}{60} \text{ joule per cm}^2 \text{ per second} \\ &= 0.067 \text{ watt per cm}^2. \end{aligned}$$

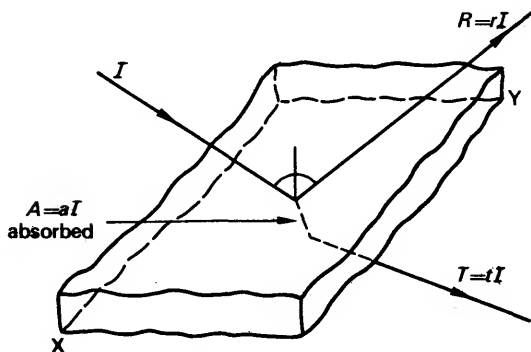


FIG. 13.17. Reflection, transmission, and absorption.

Some of this energy is reflected by the glass (R), some is absorbed (A), and some is transmitted (T). The total energy transmitted, absorbed and reflected per cm^2 per second is equal to the energy falling on the body over the same area and in the same time:

$$T + A + R = I.$$

If we denote by t , a , and r , the fractions of energy which are respectively transmitted, absorbed, and reflected by the body, then

$$tI + aI + rI = I$$

or

$$t + a + r = 1 \quad (6)$$

This equation expresses common knowledge: if a body is transparent ($t \rightarrow 1$), it is not opaque, and it is not a good reflector ($a \rightarrow 0$, $r \rightarrow 0$). But also, if the body is a good absorber of radiation ($a \rightarrow 1$), it is not transparent, and its surface is dull ($t \rightarrow 0$, $r \rightarrow 0$). And if it is a good reflector ($r \rightarrow 1$), it is neither transparent nor a good absorber ($t \rightarrow 0$, $a \rightarrow 0$). The term opaque, as commonly used, simply means not transparent; we see that it does not necessarily mean absorbent.

Equation (6), as we have written it above, is over-simplified. For a body may transmit some wavelengths (colours, if visible) and absorb or reflect others. If we now let I denote the intensity of radiation of a particular wavelength λ , then by repeating the argument we get

$$t_\lambda + a_\lambda + r_\lambda = 1 \quad (7)$$

where the coefficients t_λ , etc., all refer to the wavelength λ .

The truth of equation (7) is well shown by the metal gold, which reflects yellow light better than other colours. In thin films, gold is partly transparent, and the light which it transmits is green. Green is the colour complementary to yellow; gold removes the yellow from white light by reflection, and passes on the rest by transmission.

Radiation and Absorption

We have already pointed out that black surfaces are good absorbers and radiators of heat, and that polished surfaces are bad absorbers and radiators. This can be demonstrated by the apparatus in Fig. 13.18, in which is a cubical metal tank whose sides have a variety of finishes: dull black, dull white, highly polished. It contains boiling water, and, therefore, has a constant temperature. Facing it is a thermopile, P ,

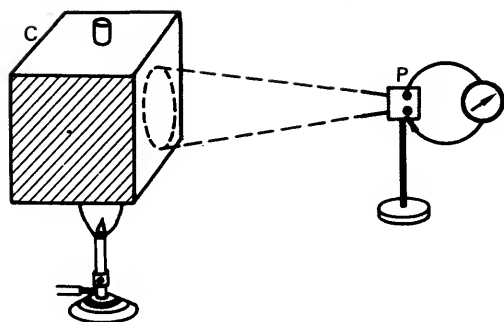


FIG. 13.18. Comparing radiators.

which is fitted with the blackened conical mouth-piece described on p. 347.

Provided that the face of the cube occupies the whole field of view of the cone, its distance from the thermopile does not matter (p. 347). The galvanometer deflection is greatest when the thermopile is facing the dull black surface of the cube, and least when it is facing the highly polished surface. The highly polished surface is therefore the worst radiator of all, and the dull black is the best.

This experiment was first done by Leslie in 1804. There were no thermopiles in those days, and Leslie detected the radiant heat with an instrument depending on the expansion of air, which we shall not describe. The tank with different surfaces is called *Leslie's cube*.

Leslie's cube can also be used in an experiment to compare the absorbing properties of surfaces, due to Ritchie (1833). A modern version of it is shown in Fig. 13.19. The cube C, full of boiling water, is placed between two copper plates, A, B, of which A is blackened and B is polished. The temperature difference between A and B is measured by making each of them one element in a thermo junction: they are joined by a constantan wire, XY, and connected to a galvanometer, by copper wires, AE, DB. If A is hotter than B, the junction, X, is hotter than the junction, Y, and a current flows through the galvanometer in one direction. If B is hotter than A, the current is reversed.

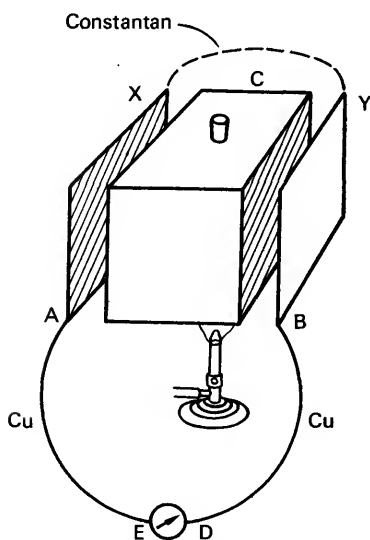


FIG. 13.19. Comparing absorbers.

The most suitable type of Leslie's cube is one which has two opposite faces similar—say grey—and the other two opposite faces very dissimilar—one black, one polished. At first the plates A, B are set

opposite similar faces. The blackened plate, A, then becomes the hotter, showing that it is the better absorber.

The cube is now turned so that the blackened plate, A, is opposite the polished face of the cube, while the polished plate, B, is opposite the blackened face of the cube. The galvanometer then shows no deflection; the plates thus reach the same temperature. It follows that the good radiating property of the blackened face of the cube, and the bad absorbing property of the polished plate, are just compensated by the good absorbing property of the blackened plate, and the bad radiating property of the polished face of the cube.

The Thermos Flask

A thermos flask—sometimes called a *Dewar flask* after its inventor (c. 1894)—is a device for reducing the transfer of heat to a minimum. It consists of a double walled glass vessel, as shown in Fig. 13.20; the space between the walls is exhausted to as high a vacuum as possible, and the insides of the walls are silvered. Heat therefore passes very slowly from the outer wall to the inner by radiation. If the vacuum is good, convection is almost inhibited—the goodness of the vacuum determines the goodness of the flask. Conduction through the glass is slight, because the conduction paths are long. In a good flask, the main cause of heat loss is conduction through the cork.

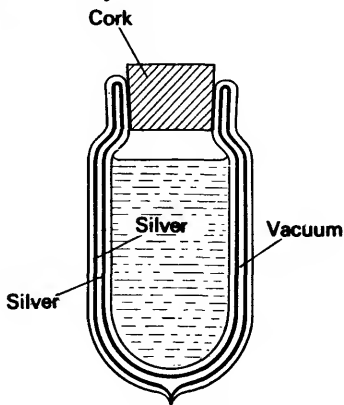


FIG. 13.20. A thermos flask.

The Black Body

The experiments described before lead us to the idea of a perfectly black body; one which absorbs all the radiation that falls upon it, and reflects and transmits none. The experiments also lead us to suppose that such a body would be the best possible radiator.

A perfectly black body can be very nearly realized—a good one can be made in half a minute, simply by punching a small hole in the lid of an empty tin. The hole looks almost black, although the shining tin is a good reflector. The hole looks black because the light which enters through it is reflected many times round the walls of the tin, before it meets the hole again (Fig. 13.21). At each reflection, about 80 per cent of the light energy is reflected, and 20 per cent is absorbed. After two reflections, 64 per cent of the

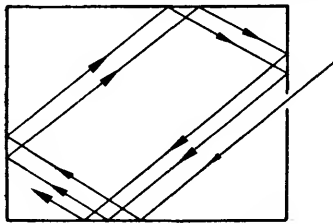


FIG. 13.21. Multiple reflections make a black body.

original light goes on to be reflected a third time; 36 per cent has been absorbed. After ten reflections, the fraction of the original energy which has been absorbed is 0.8^{10} , or 0.1.

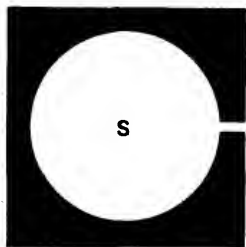


FIG. 13.22. A black body.

Any space which is almost wholly enclosed approximates to a black body. And, since a good absorber is also a good radiator, an almost closed space is the best radiator we can find.

A form of black-body which is used in radiation measurements is shown in Fig. 13.22. It consists of a porcelain sphere, S, with a small hole in it. The inside is blackened with soot to make it as good a radiator and as bad a reflector as possible. (The effect of multiple reflections is then to convert the body from nearly black to very nearly black indeed.)

The sphere is surrounded by a high-temperature bath of, for example, molten salt (the melting-point of common salt is 801°C).

The deepest recesses of a coal or wood fire are black bodies. Anyone who has looked into a fire knows that the deepest parts of it look brightest—they are radiating most power. Anyone who has looked into a fire also knows that, in the hottest part, no detail of the coals or wood can be seen. That is to say, the radiation from an almost enclosed space is uniform; its character does not vary with the nature of the surfaces of the space. This is so because the radiation coming out from any area is made up partly of the radiation emitted by that area, and partly of the radiation from other areas, reflected at the area in question. If the surface of the area is a good radiator, it is a bad reflector, and vice-versa. And if the hole in the body is small, the radiations from every area inside it are well mixed by reflection before they can escape; the intensity and quality of the radiation escaping thus does not depend on the particular surface from which it escapes.

When we speak of the *quality* of radiation we mean the relative intensities of the different wavelengths that it comprises; the proportion of red to blue, for example. The quality of the radiation from a perfectly black body depends only on its temperature. When the body is made hotter, its radiation becomes not only more intense, but also more nearly white; the proportion of blue to red in it increases. Because its quality is determined only by its temperature, black-body radiation is sometimes called 'temperature radiation'.

Properties of Temperature Radiation

The quality of the radiation from a black body was examined by Lummer and Pringsheim in 1899. They used a black body represented by B in Fig. 13.23 and measured its temperature with a thermocouple; they took it to 2000°C . To measure the intensities of the various wavelengths, Lummer and Pringsheim used an infra-red spectrometer and a bolometer (p. 346) consisting of a single platinum strip.

The results of experiments such as these are shown in Fig. 13.24 (a). Each curve gives the relative intensities of the different wavelengths,

for a given temperature of the body. The curves show that, as the temperature rises, the intensity of every wavelength increases, but the intensities of the shorter wave lengths increase more rapidly. Thus the radiation becomes, as we have already observed, less red, that is to say, more nearly white. The curve for sunlight has its peak at about 5000 A.U., in the visible green; from the position of this peak we conclude that the surface temperature of the sun is about 6000 K. Stars which are hotter than the sun, such as Sirius and Vega, look blue, not white.

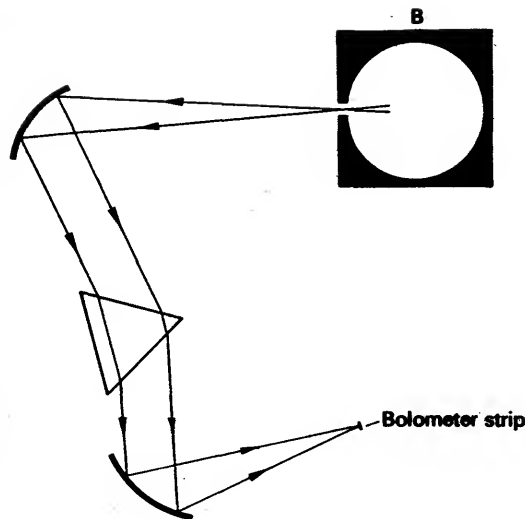


FIG. 13.23. Lummer and Pringsheim's apparatus for study of black body radiation (diagrammatic).

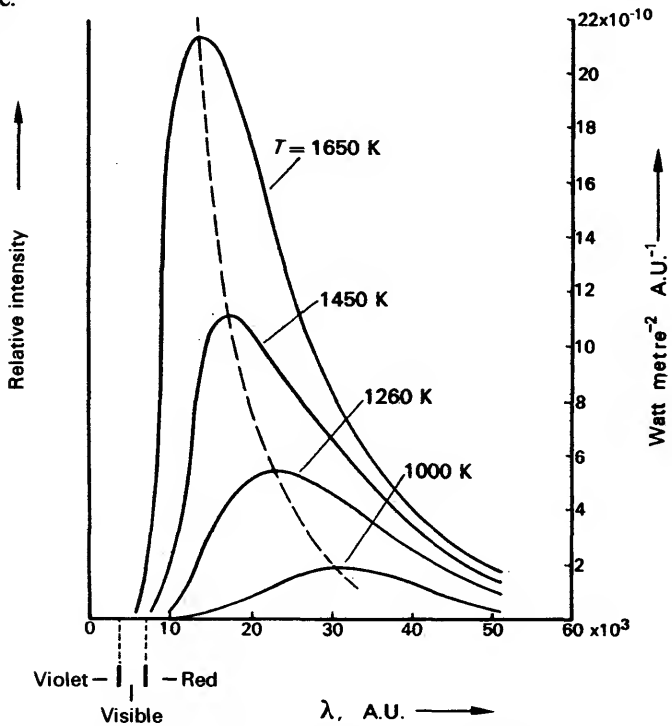


FIG. 13.24 (a). Distribution of intensity in black-body radiation.

because the peaks of their radiation curves lie further towards the visible blue than does the peak of sunlight.

The actual intensities of the radiations are shown on the right of the graph in Fig. 13.24 (a). To speak of the intensity of a single wavelength is meaningless because there is an infinite number of wavelengths, but the total intensity of the radiation is finite. The slit of the spectrometer always gathers a band of wavelengths—the narrower the slit the narrower the band—and we always speak of the intensity of a given band. We express it as follows ('s' represents 'second'):

$$\text{energy radiated } \text{m}^{-2} \text{ s}^{-1}, \text{ in band } \lambda \text{ to } \lambda + \delta\lambda = E_{\lambda} \delta\lambda. \quad (8)$$

The quantity E_{λ} is called the *emissive power* of a black body for the wavelength λ and at the given temperature; its definition follows from equation (8):

$$E_{\lambda} = \frac{\text{energy radiated } \text{m}^{-2} \text{ s}^{-1}, \text{ in band } \lambda \text{ to } \lambda + \delta\lambda}{\text{bandwidth, } \delta\lambda}.$$

The expression 'energy per second' can be replaced by the word 'power', whose unit is the watt. Thus

$$E_{\lambda} = \frac{\text{power radiated } \text{m}^{-2} \text{ in band } \lambda, \lambda + \delta\lambda}{\delta\lambda}.$$

In the figure E_{λ} is expressed in watt per m^2 per Angstrom unit.

SI units may be 'watt per metre² per nanometre (10^{-9} m)'.

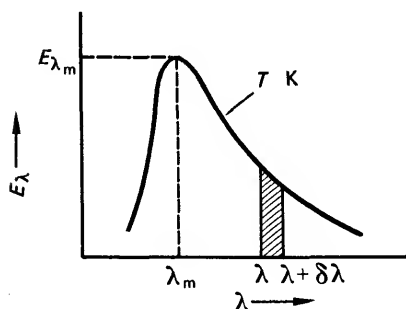


FIG. 13.24 (b). Definition of E , λ_m , and E_{λ_m} .

The quantity $E_{\lambda} \delta\lambda$ in equation (8) is the area beneath the radiation curve between the wavelengths λ and $\lambda + \delta\lambda$ (Fig. 13.24 (b)). Thus the energy radiated per cm^2 per second between those wavelengths is proportional to that area. Similarly the total radiation emitted per cm^2 per second over all wavelengths is proportional to the area under the whole curve.

Laws of Black Body Radiation

The curves of Fig. 13.24 (a) can be explained only by the quantum theory of radiation, which is outside our scope. Both theory and experiment lead to three generalizations, which together describe well the properties of black-body radiation:

- (i) If λ_m is the wavelength of the peak of the curve for T K, then

$$\lambda_m T = \text{constant} \quad (9)$$

The value of the constant is 2.9×10^{-3} m K. In Fig. 13.24 (a) the dotted line is the locus of the peaks of the curves for different temperatures.

(ii) If E_{λ_m} is the height of the peak of the curve for the temperature T K, then

$$E_{\lambda_m} \propto T^5 \quad (10)$$

The relationships (10) and (9) are particular cases of a general law given by Wien in 1894; (9) is sometimes called *Wien's displacement law*.

(iii) If E is the total energy radiated per metre² per second at a temperature T , represented by the area under the curve, then

$$E = \sigma T^4,$$

where σ is a constant. This result is called **Stefan's law**, and the constant σ is called *Stefan's constant*. Its value is

$$\sigma = 5.7 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}.$$

Prévost's Theory of Exchanges

In 1792 Prévost applied the idea of dynamic equilibrium to radiation. He asserted that a body radiates heat at a rate which depends only on its surface and its temperature, and that it absorbs heat at a rate depending on its surface and the temperature of its surroundings. When the temperature of a body is constant, the body is losing heat by radiation, and gaining it by absorption, at equal rates.

It is easy to think of experiments which seem to support Prévost's theory, and the reader will certainly grasp the general idea of it if he imagines hot pies and cold ice-creams put into the same cupboard. But in such experiments it is difficult to get rid of the possibility of convection. Let us rather take an old-fashioned, high vacuum, electric lamp, and put it in a can of water (Fig. 13.25 (a)). We can find the temperature of the lamp's filament by measuring its resistance. We find that, whatever the temperature of the water, the filament comes to that temperature, if we leave it long enough. When the water is cooler than the filament, the filament cools down; when the water is hotter, the filament warms up.

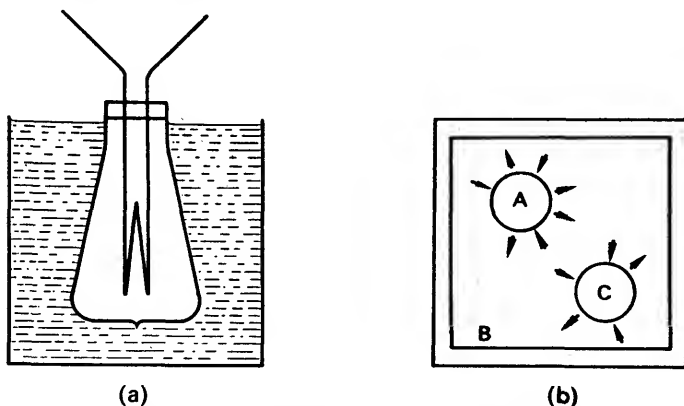


FIG. 13.25. Illustrating Prévost's theory of exchanges.

In the abstract language of theoretical physics, Prévost's theory is easy enough to discuss. If a hot body A (Fig. 13.25 (b)) is placed in an evacuated enclosure B, at a lower temperature than A, then A cools until it reaches the temperature of B. If a body C, cooler than B, is put in B, then C warms up to the temperature of B. We conclude that radiation from B falls on C, and therefore also on A, even though A

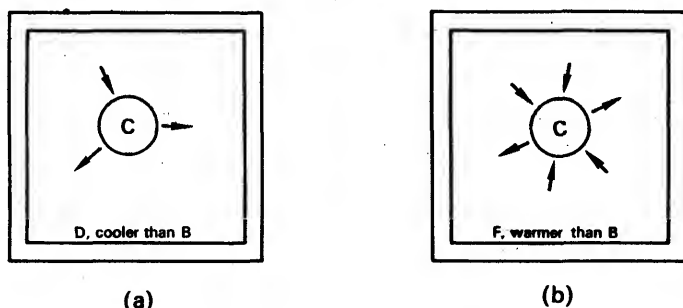


FIG. 13.26. Illustrating Prévost's theory.

is at a higher temperature. Thus A and C each come to equilibrium at the temperature of B when each is absorbing and emitting radiation at equal rates.

Now let us suppose that, after it has reached equilibrium with B, one of the bodies, say C, is transferred from B to a cooler evacuated enclosure D (Fig. 13.26 (a)). It loses heat and cools to the temperature of D. Therefore it is radiating heat. But if C is transferred from B to a warmer enclosure F, then C gains heat and warms up to the temperature of F (Fig. 13.26 (b)). It seems unreasonable to suppose that C stops radiating when it is transferred to F; it is more reasonable to suppose that it goes on radiating but, while it is cooler than F, it absorbs more than it radiates.

Emissivity

Let us consider a body B, in equilibrium with an enclosure A, at a temperature T K (Fig. 13.27). If the body is perfectly black, it emits radiation characteristics of the temperature T ; let us write the total intensity of this radiation over all wavelengths as E watts/m². Since the body is in equilibrium with the enclosure, it is absorbing as much as it radiates. And since it absorbs all the radiation that falls upon it, the energy falling on it per cm² per second must be equal to E . This conclusion need not surprise us, since the enclosure A is full of black body radiation characteristic of its temperature T .

Now let us consider, in the same enclosure, a body C which is not black. On each square metre of the body's surface, E watts of radiation fall (Fig. 13.27). Of this, let us suppose that the body absorbs a fraction a , that is to say, it absorbs aE watts per m². We may call a the total absorption factor of the body C, 'total' because it refers to the total radiation. The radiation which the body does not absorb, $(1 - a)E$, it reflects or transmits.

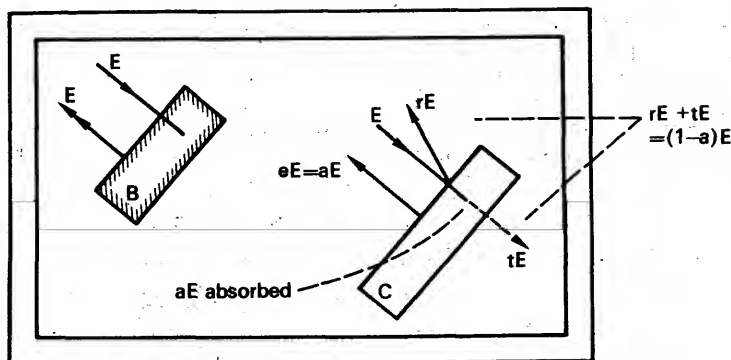


FIG. 13.27. Equilibria in an enclosure.

Thus:
$$\left. \begin{array}{l} \text{power reflected or} \\ \text{transmitted/m}^2 \end{array} \right\} = E - aE.$$

For equilibrium, the total power leaving the body per m^2 must be equal to the total power falling upon it, $E \text{ W/m}^2$. The power emitted by the body, which must be added to that reflected and transmitted, is therefore:

$$\text{total power radiated/m}^2 = aE \quad (12)$$

The ratio of the total power radiated per m^2 by a given body, to that emitted by a black body at the same temperature, is called the *total emissivity* of the given body. Hence, by equation (12),

$$e = \frac{aE}{E} = a.$$

We have therefore shown that the total emissivity of a body is equal to its total absorption factor.

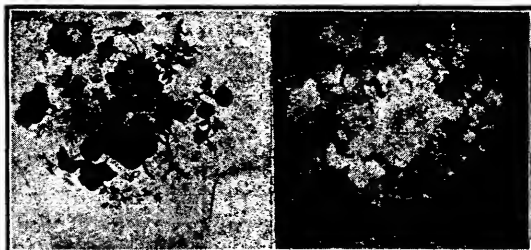
This is a formal expression of the results of Ritchie's experiment (p. 350). If we combine it with Stefan's law, we find that the total energy E radiated per m^2 per second by a body of emissivity e at a temperature $T \text{ K}$ is

$$E = eE = e\sigma T^4.$$

Spectral Emissivity; Kirchhoff's Law

Most bodies are coloured; they transmit or reflect some wavelengths better than others. We have already seen that they must absorb these wavelengths weakly; we now see that, because they absorb them weakly, they must also

FIG. 13.28. Photographs showing how a piece of incandescent decorated crockery appears (a) by reflected light and (b) by its own emitted light.



with the temperature. A body for which e_λ is the same for all wavelengths, but is less than unity, is said to be 'grey'.

Absorption by Gases

An experiment which shows that, if a body radiates a given wavelength strongly, it also absorbs that wavelength strongly, can be made with sodium vapour. A sodium vapour lamp runs at about 220°C ; compared with the sun, or even an arc-lamp, it is cool. The experiment consists of passing sunlight or arc-light through a spectroscope, and observing its continuous spectrum. The sodium lamp is then placed in the path of the light, and a black line appears in the yellow. If the white light is now cut off, the line which looked black comes up brightly—it is the sodium yellow line.

The process of absorption by sodium vapour—or any other gas—is not, however, the same as the process of absorption by a solid. When a solid absorbs radiation, it turns it into heat—into the random kinetic energy of its molecules. It then re-radiates it in all wavelengths, but mostly in very long ones, because the solid is cool. When a vapour absorbs light of its characteristic wavelength, however, its atoms are excited; they then re-radiate the absorbed energy, in the same wavelength (5893 A.U. for sodium). But they re-radiate it in all directions, and therefore less of it passes on in the original direction than before (Fig. 13.30). Thus the yellow component of the original beam is weakened, but the yellow light radiated sideways by the sodium is strengthened. The sideways strengthening is hard to detect, but it was shown by R. W. Wood in 1906. He used mercury vapour instead of sodium. The phenomenon is called *optical resonance*, by analogy with resonance in sound.

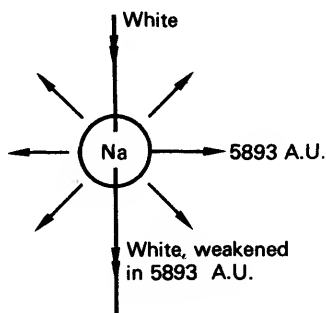


FIG. 13.30. Absorption by sodium vapour.

CONVECTION

Liquids—except mercury, which is a molten metal—are bad conductors of heat. If we hold a test-tube full of water by the bottom, we can boil the water near the top in a Bunsen flame, without any discomfort. But if we hold the tube at the top, and heat it at the bottom, then the top becomes unbearably hot long before the water boils. The heat is brought to the top by *convection*; the warm water at the bottom expands, becomes less dense, and rises; the cold water sinks to take its place. If we heat a beaker of water at one side, and drop in a crystal of potassium permanganate, we can see the currents of hot water rising, and cold descending. Central-heating systems rely on convection to bring hot water from the boiler, in the basement, to the so-called radiators, and to take the cooler water back to the boiler.

The radiators of a central-heating system are wrongly named; they are convectors. They warm the air around them, which rises, and gives way to cooler air from cooler parts of the room. Gases are even worse conductors of heat than liquids, and for most practical purposes we can neglect conduction through them altogether. Woollen clothes keep us warm because they contain pockets of air, which hardly conduct at all, and cannot convect because they cannot move. The wool fibres themselves are much better conductors than the air they imprison.

Convection by air is important in ventilation: the fire in a room maintains a draught of hot air up the chimney, and cool fresh air from outside comes in under the door. The draught also helps to keep the fire supplied with oxygen; factory chimneys are made tall to stimulate convection and increase the draught.

Forced and Free (Natural) Convection

A gas or a liquid may carry away heat from a hot body by convection. If the flow of liquid or gas is simply due to its being heated by the body, and hence rising, the convection is said to be *free*, or natural. But if the gas or liquid is flowing in a stream maintained by some other means, then the convection is said to be *forced*. Thus cooling one's porridge in the obvious way is an example of forced convection; it causes a more rapid loss of heat than does natural convection.

Critical Diameter of Pipes

Hot-water and steam pipes are often lagged with asbestos to reduce the loss of heat from them. The temperature drop across the lagging makes the outside cooler than the pipe, and so, by Newton's law, tends to reduce the rate at which heat escapes from it. However, the lagging increases the outside diameter of the whole, and so increases its area of contact with the atmosphere. The increase in area tends to make convection more vigorous, by enabling the pipe to heat a greater mass of air. If the diameter of the pipe is small, the increase in area may more than offset the reduction temperature of the outside, and so increase the rate of heat loss. Thus there is a critical diameter of pipe; if the diameter is less than the critical value, the pipe should not be lagged. The critical diameter depends on many factors, but is commonly of the order of 1 cm.

The Greenhouse

A greenhouse keeps plants warm by inhibiting convection. The glass allows radiant heat to reach the plants from the sun, but prevents the warm air in the greenhouse from escaping. In winter, when there is little sunshine, the heat is provided by hot water pipes. In summer the temperature is regulated by opening or closing the roof and windows, and so adjusting the loss of heat by convection.

EXERCISES 13

Conduction

1. Define the *thermal conductivity* of a substance. Describe how the thermal conductivity of a metal may be measured, pointing out the sources of error in the experiment.

A large hot-water tank has four steel legs in the form of cylindrical rods 2.5 cm in diameter and 15 cm long. The lower ends of the legs are in good thermal contact with the floor, which is at 20°C, and their upper ends can be taken to be at the temperature of the water in the tank. The tank and the legs are well lagged so that the only heat loss is through the legs. It is found that 22 watts are needed to maintain the tank at 60°C. What is the thermal conductivity of steel? When a sheet of asbestos 1.5 mm thick is placed between the lower end of each leg and the floor only 5 watts are needed to maintain the tank at 60°C. What is the thermal conductivity of asbestos? (*O. & C.*)

2. Define *thermal conductivity* and state a unit in which it is expressed.

Explain why, in an experiment to determine the thermal conductivity of copper using a Searle's arrangement, it is necessary (*a*) that the bar should be thick, of uniform cross-section and have its sides well lagged, (*b*) that the temperatures used in the calculation should be the steady values finally registered by the thermometers.

Straight metal bars *X* and *Y* of circular section and equal in length are joined end to end. The thermal conductivity of the material of *X* is twice that of the material of *Y*, and the uniform diameter of *X* is twice that of *Y*. The exposed ends of *X* and *Y* are maintained at 100°C and 0°C, respectively and the sides of the bars are ideally lagged. Ignoring the distortion of the heat flow at the junction, sketch a graph to illustrate how the temperature varies between the ends of the composite bar when conditions are steady. Explain the features of the graph and calculate the steady temperature of the junction. (*N.*)

3. Give a critical account of an experiment to determine the thermal conductivity of a material of low thermal conductivity such as cork. Why is it that most cellular materials, such as cotton wool, felt, etc., all have approximately the same thermal conductivity?

One face of a sheet of cork, 3 mm thick, is placed in contact with one face of a sheet of glass 5 mm thick, both sheets being 20 cm square. The outer faces of this square composite sheet are maintained at 100°C and 20°C, the glass being at the higher mean temperature. Find (*a*) the temperature of the glass-cork interface, and (*b*) the rate at which heat is conducted across the sheet, neglecting edge effects.

[Thermal conductivity of cork = $6.3 \times 10^{-2} \text{ W m}^{-1} \text{ K}^{-1}$, thermal conductivity of glass = $7.14 \times 10^{-1} \text{ W m}^{-1} \text{ K}^{-1}$.] (*O. & C.*)

4. Define *coefficient of thermal conductivity*. Describe a method of measuring this coefficient for a metal.

Assuming that the thermal insulation provided by a woollen glove is equivalent to a layer of quiescent air 3 mm thick, determine the heat loss per minute from a man's hand, surface area 200 cm² on a winter's day when the atmospheric air temperature is -3°C. The skin temperature is to be taken as 34°C and the thermal conductivity of air as $24 \times 10^{-3} \text{ W m}^{-1} \text{ K}^{-1}$. (*L.*)

5. Define *thermal conductivity*. Describe in detail a method of determining the thermal conductivity of cork in the form of a thin sheet.

The base and the vertical walls of an open thin-walled metal tank, filled with water maintained at 35°C, are lagged with a layer of cork of superficial area 2.00 m²

and 1.00 cm thick and the water surface is exposed. Heat is supplied electrically to the water at the rate of 250 watts. Find the mass of water that will evaporate per day, if the outside surface of the cork is at 15°C . [Assume that the thermal conductivity of cork is $5.0 \times 10^{-2} \text{ W m}^{-1} \text{ K}^{-1}$ and that the latent heat of vaporization of water at 35°C is 2520 J g^{-1} .] (L.)

6. Explain what is meant by the *coefficient of thermal conductivity* of a metal.

One end of a long uniform metal bar is heated in a steam chest and the other is kept cool by a current of water. Draw sketch graphs to show the variation of temperature along the bar when the steady state has been attained (a) when the bar is lagged so that no heat escapes from the sides, (b) when the bar is exposed to the air. Explain the shape of the graph in each case.

The surface temperatures of the glass in a window are 20°C for the side facing the room and 5°C for the outside. Compare the rate of flow of heat through (i) a window consisting of a single sheet of glass 5.0 mm thick, and (ii) a double-glazed window of the same area consisting of two sheets of glass each 2.5 mm thick separated by a layer of still air 5.0 mm thick. It may be assumed that the steady state has been attained.

[Use the following values of coefficient of thermal conductivity: glass: $1.0 \text{ W m}^{-1} \text{ K}^{-1}$; air: $2.5 \times 10^{-2} \text{ W m}^{-1} \text{ K}^{-1}$.] (C.)

7. Describe the construction of a Dewar (vacuum) vessel and explain the physical features which result in a reduction to a minimum of the heat exchange between the interior and exterior. Explain why such a vessel is equally suitable for thermally isolating a cold or a hot body.

A copper sphere of radius 0.5 cm is suspended in an evacuated enclosure by a copper wire of diameter 0.01 cm and length 3 cm. An insulated electrical heating coil in good thermal contact with the sphere is connected through the wall of the enclosure by two copper leads of negligible resistance each of diameter 0.02 cm and length 5 cm. What rate of heating in the coil is required to maintain the sphere at a temperature 50°C above that of the surroundings assuming that heat is lost only by conduction along the supports and along the electrical leads? When a steady state has been reached, the coil is disconnected from the electrical supply and the initial rate of fall of temperature of the sphere is found to be 0.013°C per second. Calculate the specific heat of copper, assuming that the electrical leads are still kept at the temperature of the surroundings at the points where they pass through the wall of the enclosure, that the temperature gradient in the sphere is negligible and given that the thermal capacity of the heating coil is equal to that of 1.5 g of copper. [Density of copper = 9.0 g cm^{-3} . Thermal conductivity of copper = $380 \text{ W m}^{-1} \text{ K}^{-1}$.] (O. & C.)

8. Define *thermal conductivity*.

A 'cold probe', i.e. an instrument to produce a low temperature at its extremity, consists of a solid copper rod 1 mm in diameter attached axially to a well-lagged

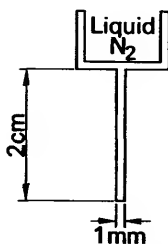


FIG. 13A.

copper reservoir which holds boiling liquid nitrogen, temperature 78 K (see Fig. 13A). The distance from the bottom of the reservoir to the tip of the probe is 2.0 cm. The curved surface of the rod is coated with a non-conducting material. Assuming that no heat can reach the copper except through its flat tip, calculate the maximum rate at which heat can be accepted there if the temperature is not to rise above -10°C .

(Thermal conductivity of copper = $385 \text{ W m}^{-1} \text{ K}^{-1}$.)

If you were given such a probe and told to use it to determine the mean conductivity of copper over this temperature range, how would you proceed? Describe the kind of apparatus you would use and specify (a) the quantities you would need to measure, (b) any data you would need to know. (O. & C.)

Radiation

9. A hot body, such as a wire heated by an electric current, can lose energy to its surroundings by various processes. Outline the nature of each of these processes.

A black body of temperature t is situated in a blackened enclosure maintained at a temperature of 10°C . When $t = 30^{\circ}\text{C}$ the net rate of loss of energy from the body is equal to 10 watts. What will the rate become when $t = 50^{\circ}\text{C}$ if the energy exchange takes place solely by the process of radiation? What percentage error is there in the answer obtained by basing the solution on Newton's law of cooling? (C.)

10. Explain what is meant by a *black body*. How do the total energy radiated by a black body and its distribution among the wavelengths in the spectrum depend upon the temperature of the radiator?

Describe the structure of an optical pyrometer and explain how it is used to measure the temperature of a furnace. (L.)

11. Explain what is meant by *black body radiation* and how it can be obtained in practice.

Give an account of Prévost's theory of exchanges and show how it can be used in conjunction with Stefan's law to obtain an expression for the net rate of loss of heat by a black body cooling in an evacuated enclosure.

Sketch the curves relating intensity of radiation and wavelength of radiation from a black body, for three different temperatures. (L.)

12. Explain what is meant by *Stefan's constant*. Defining any symbols used.

A sphere of radius 2.00 cm with a black surface is cooled and then suspended in a large evacuated enclosure the black walls of which are maintained at 27°C . If the rate of change of thermal energy of the sphere is 1.848 J s^{-1} when its temperature is -73°C , calculate a value for *Stefan's constant*. (N.)

13. What is Prévost's Theory of Exchanges? Describe some phenomenon of theoretical or practical importance to which it applies.

A metal sphere of 1 cm diameter, whose surface acts as a black body, is placed at the focus of a concave mirror with aperture of diameter 60 cm directed towards the sun. If the solar radiation falling normally on the earth is at the rate of $0.14 \text{ watt cm}^{-2}$, Stefan's constant is taken as $6 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ and the mean temperature of the surroundings is 27°C , calculate the maximum temperature which the sphere could theoretically attain, stating any assumptions you make. (O. & C.)

14. State Newton's law of cooling and Stefan's fourth power law. Describe an experiment to test the validity of one of these laws.

A sphere of copper cools at the rate of $10 \text{ deg C min}^{-1}$ when at a temperature of 70°C in an enclosure at 20°C . Calculate its rate of cooling when its temperature

is raised to (a) 100°C , (b) 700°C , assuming the validity of Newton's law. Repeat the calculations assuming the validity of Stefan's law. Comment on your answers. (L.)

15. Explain what is meant by (a) a *black body*, (b) *black body radiation*.

State Stefan's law and draw a diagram to show how the energy is distributed against wavelength in the spectrum of a black body for two different temperatures. Indicate which temperature is the higher.

A roof measures $20\text{ m} \times 50\text{ m}$ and is blackened. If the temperature of the sun's surface is 6000 K , Stefan's constant $= 5.72 \times 10^{-8}\text{ W m}^{-2}\text{ K}^{-4}$, the radius of the sun is $7.5 \times 10^8\text{ m}$ and the distance of the sun from the earth is $1.5 \times 10^{11}\text{ m}$, calculate how much solar energy is incident on the roof per minute, assuming that half is lost in passing through the earth's atmosphere, the roof being normal to the sun's rays. (O. & C.)

16. What is *black body radiation*?

Using the same axes sketch graphs, one in each instance, to illustrate the distribution of energy in the spectrum of radiation emanating from (a) a black body at 1000 K , (b) a black body at 2000 K and (c) a source other than a black body at 1000 K . Point out any special features of the graphs.

Indicate briefly how the relative intensities needed to draw one of these graphs could be determined. (N.)

17. How can the temperature of a furnace be determined from observations on the radiation emitted?

Calculate the apparent temperature of the sun from the following information :

Sun's radius: $7.04 \times 10^5\text{ km}$.

Distance from earth: $14.72 \times 10^7\text{ km}$.

Solar constant: $0.14\text{ watt per cm}^2$.

Stefan's constant: $5.7 \times 10^{-8}\text{ W m}^{-2}\text{ K}^{-4}$. (N.)

18. Describe one experiment to show that a polished metal surface is a poor absorber of heat, and one experiment to show that such a surface reflects a high proportion of a beam of light falling on it. Briefly compare heat and light radiations from the standpoint of (a) velocity, (b) effect at a distance, (c) simple refraction, (d) transmission through material substances. (N.)

19. A block of metal is heated and (a) exposed to ordinary atmospheric conditions, or (b) placed in a high vacuum. State concisely the factors that govern the rate at which its temperature falls under conditions (a) and (b). Energy is supplied at the rate of 165 watts to a closed cylindrical canister 5 cm in radius and 15 cm high, filled with water and exposed to the air of the room, which is at 15°C . It is found that the temperature of the water remains steady at 80°C . Find the rate of heat loss per unit area of the vessel per deg C excess temperature. Estimate also the fall of temperature in a minute, when the energy supply is shut off. Neglect the weight of the canister itself. (L.)

20. Describe an experiment to show (a) that the spectrum of an incandescent solid includes both visible and invisible radiations, (b) how the fraction of incident radiation transmitted by glass depends on the temperature of the source of the radiation.

The sun's rays are focussed by a concave mirror of diameter 12 cm fixed with its axis towards the sun on to a copper calorimeter, where they are absorbed. If the thermal capacity of the calorimeter and its contents is $247.8\text{ joules per deg C}$ and the temperature rises 8°C in 2 min , calculate the heat received in 1 min by a square metre of the earth's surface when the rays are incident normally. (N.)

21. Give an account of Stefan's law of radiation, explaining the character of the radiating body to which it applies and how such a body can be experimentally realized.

If each square cm of the sun's surface radiates energy at the rate of $6.3 \times 10^3 \text{ J s}^{-1} \text{ cm}^{-2}$ and Stefan's constant is $5.7 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$, calculate the temperature of the sun's surface in degrees centigrade, assuming Stefan's law applies to the radiation. (L.)

22. Explain what is meant by *black body radiation* and how it can be obtained in practice.

Give an account of Prévost's theory of exchanges and show how it can be used in conjunction with Stefan's law to obtain an expression for the net rate of loss of heat by a black body cooling in an evacuated enclosure.

Sketch the curves relating intensity of radiation and wavelength of radiation from a black body, for *three* different temperatures. (L.)

chapter fourteen

Thermometry and Pyrometry

REALIZATION OF TEMPERATURE SCALE

IN Chapter 8 we discussed the general idea of a temperature scale. To establish such a scale we need:

- (i) some physical property of a substance—such as the volume of a particular liquid—which increases continuously with increasing hotness, but is constant at constant hotness;
- (ii) two standard degrees of hotness—the fixed points (ice and steam)—which can be accurately reproduced.

The Fixed Points and Fundamental Interval

The temperature of melting ice and saturated steam are chosen as the fixed points on the Celsius temperature scale. We have seen that the atmospheric pressure must be specified in defining the steam point, but this is not necessary in defining the ice point, because the melting-point of ice changes very little with pressure (p. 296). On the other hand impurities in water do not affect the temperature of saturated steam, but impurities in ice do affect its melting-point (p. 221); the ice used in realizing the lower fixed point must therefore be prepared from pure water.

Then, if P is the chosen temperature-measuring quantity, its values P_0 at the ice point, and P_{100} at the steam point determine the fundamental interval of the scale: $P_{100} - P_0$. And the temperature θ_p on the P -scale, which corresponds to a value P_θ of P is, *by definition*,

$$\theta_p = \frac{P_\theta - P_0}{P_{100} - P_0} \times 100.$$

On the thermodynamic scale, the triple point of water (p. 319) is chosen as one fixed point and is defined as 273.16 K. The other fixed point is the absolute zero (see p. 190).

The Thermometric Substance and Property

Most thermometers are of the liquid-in-glass type, because it is simple and cheap; they contain either mercury or alcohol.

The mercury and alcohol scales agree fairly well with one another, and with either of the gas scales. The gas scales depend on the change of volume at constant pressure and of pressure at constant volume (p. 225). In practice the constant volume scale is always used, because a change of pressure is easier to measure accurately than a change of volume.

The mercury scale agrees better with the gas scales than does the

alcohol scale. However, even the best mercury thermometers disagree slightly amongst themselves. The discrepancies may arise because the bores of the tubes are not uniform, or the mercury is impure, or the glass is not homogeneous.

In most accurate work, therefore, temperatures are measured by the changes in pressure of a gas at constant volume. At pressures of the order of one atmosphere, different gases give slightly different temperature scales, because none of them obeys the gas laws perfectly. But as the pressure is reduced, the gases approach closely to the ideal, and their temperature scales come together. By observing the departure of a gas from Boyle's law at moderate pressures it is possible to allow for its departure from the ideal; temperatures measured with the gas in a constant volume thermometer can then be converted to the values which would be given by the same thermometer if the gas were ideal.

The Constant-Volume Gas Thermometer

Fig. 14.1 shows a constant volume hydrogen thermometer, due to Chappius (1884). B is a bulb of platinum-iridium, holding the gas.

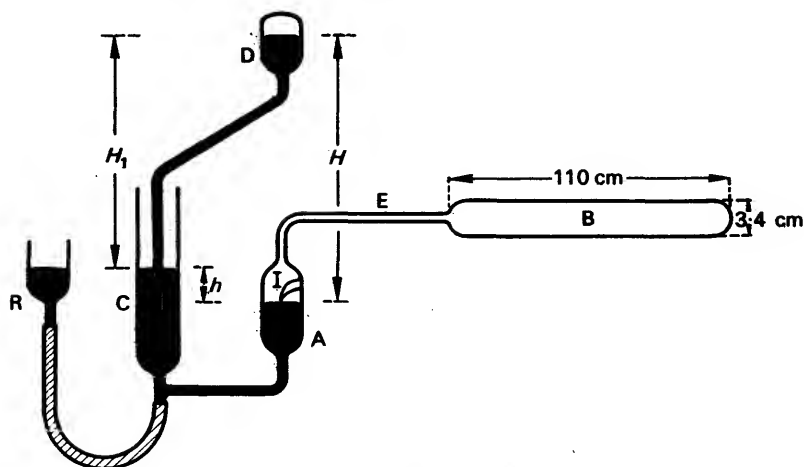


FIG. 14.1. Constant volume hydrogen thermometer (not to scale).

The volume is defined by the level of the index I in the glass tube A. The pressure is adjusted by raising or lowering the mercury reservoir R. A barometer CD is fitted directly into the pressure-measuring system; if H_1 is its height, and h the difference in level between the mercury surfaces in A and C, then the pressure H of the hydrogen, in mm mercury is

$$H = H_1 + h.$$

H is measured with a cathetometer.

The glass tubes A, C, D, all have the same diameter to prevent errors due to surface tension; and A and D are optically worked to prevent errors due to refraction (as in looking through common window-glass).

Observations made with a constant-volume gas thermometer must be corrected for the following errors:

- (i) the expansion of the bulb B;
- (ii) the temperature of the gas in the tube E and A, which lies between the temperature of B and the temperature of the room;
- (iii) the temperature of the mercury in the barometer and manometer.

The expansion of the bulb can be estimated from its coefficient of cubical expansion, by using the temperature shown by the gas thermometer. Since the expansion appears only as a small correction to the observed temperature, the uncorrected value of the temperature may be used in estimating it. The tube E is called the 'dead-space' of the thermometer. Its diameter is made small, about 0.7 mm, so that it contains only a small fraction of the total mass of gas. Its volume is known, and the temperatures at various points in it are measured with mercury thermometers. The effect of the gas in it is then allowed for in a calculation similar to that used to calculate the pressure of a gas in two bulbs at different temperatures (p. 262). Mercury thermometers may be used to measure the temperatures because the error due to the dead-space is small; any error in allowing for it is of the second order of small quantities. For the same reason, mercury thermometers may be used to measure the temperature of the manometer and barometer.

A gas thermometer is a cumbersome instrument, demanding much skill and time, and useless for measuring changing temperatures. In practice, gas thermometers are used only for calibrating electrical thermometers—resistance thermometers and thermocouples. The readings of these, when they are used to measure unknown temperatures can then be converted into temperatures on the ideal gas scale.

The International Temperature Scale

Because of the wide use of electrical thermometers, a scale of temperature based on them is used throughout the laboratories of the world. It is called the *international scale of temperature*, and is defined so that temperatures expressed on it agree, within the limits of experimental accuracy, with the same temperatures expressed on the ideal gas scale.

For the purpose of calibrating electrical thermometers, subsidiary fixed points, in addition to the fundamental fixed points of ice and steam, have been determined with the constant-volume gas thermometer. They are all measured at an atmospheric pressure of 760 mm mercury.

Their values are given, with those of the fundamental fixed points, in the following table.

FIXED POINTS OF THE INTERNATIONAL TEMPERATURE SCALE

(a) Boiling point of liquid oxygen	.	.	-182.970°C.
(b) Ice point (fundamental)	.	.	0.000°C.
(c) Steam point (fundamental)	.	.	100.000°C.
(d) Boiling-point of sulphur	.	.	444.600°C.
(e) Freezing-point of silver	.	.	960.800°C.
(f) Freezing-point of gold	.	.	1063.000°C.

The methods of interpolating between these fixed points will be described below.

Electric Thermometers

Electrical thermometers have great advantages over other types. They are more accurate than any except gas thermometers, and are quicker in action and less cumbersome than those.

The measuring element of a *thermo-electric thermometer* is the welded junction of two fine wires. It is very small in size, and can therefore measure the temperature almost at a point. It causes very little disturbance wherever it is placed, because the wires leading from it are so thin that the heat loss along them is usually negligible. It has a very small heat capacity, and can therefore follow a rapidly changing temperature. To measure such a temperature, however, the e.m.f. of the junction must be measured with a galvanometer, instead of a potentiometer, and some accuracy is then lost.

The measuring element of a *resistance thermometer* is a spiral of fine-wire. It has a greater size and heat capacity than a thermojunction, and cannot therefore measure a local or rapidly changing temperature. But, over the range from about room temperature to a few hundred degrees Centigrade, it is more accurate.

Resistance Thermometers

Resistance thermometers are usually made of platinum. The wire is wound on two strips of mica, arranged crosswise as shown in Fig.

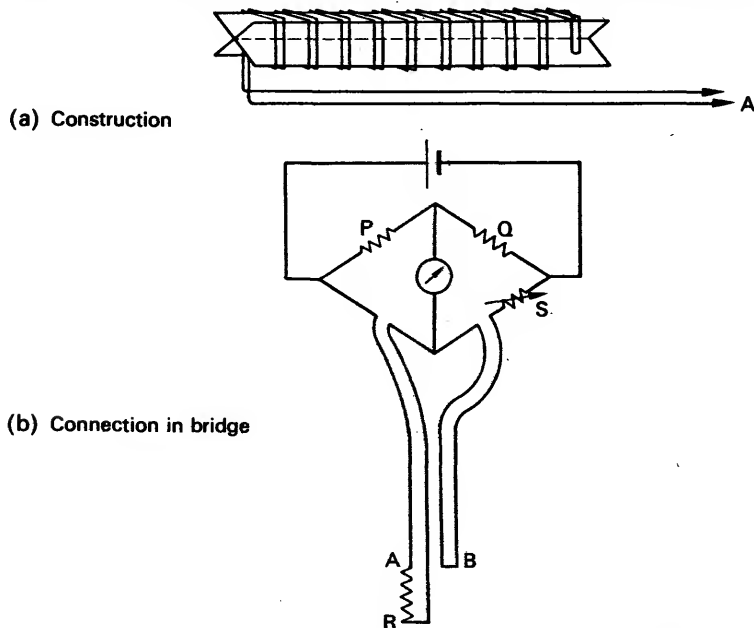


FIG. 14.2. Platinum resistance thermometer ($P = Q$, so that B compensates A ; and $S = R$).

14.2 (a). The ends of the coil are attached to a pair of leads A, for connecting them to a Wheatstone bridge. A similar pair of leads B is near to the leads from the coil, and connected in the adjacent arm of the bridge (Fig. 14.2 (b)). At the end near the coil, the pair of leads B is short-circuited. If the two pairs of leads are identical, their resistances are equal, whatever their temperature. Thus if $P = Q$ the dummy pair, B, just compensates for the pair A going to the coil; and the bridge measures the resistance of the coil alone.

The platinum resistance thermometer is used to measure temperatures on the international scale between the boiling-point of oxygen and 630°C (630°C is approximately the freezing-point of antimony, but it is not a fixed point on the scale). The platinum used in the coil must be of high purity. Its purity is judged by the increase in its resistance from the ice point to the steam point. Thus if R_0 and R_{100} are the resistances of the coil at these points, then the coil is fit to reproduce the international temperature scale if

$$\frac{R_{100}}{R_0} > 1.3910.$$

From the boiling-point of oxygen (-182.970°C) to the ice-point, the temperature θ , on the international scale, is given by the equation

$$R_{\theta} = R_0[1 + A\theta + B\theta^2 + C(\theta - 100)\theta^3] \quad (1)$$

Here R_{θ} is the resistance of the coil, and A , B , C are constants. The constants A and B are determined in a way which we shall describe shortly. When they are known the constant C can be determined from the value of R_{θ} at the boiling-point of oxygen.

From the ice-point to 630°C the temperature θ is given by

$$R_{\theta} = R_0(1 + A\theta + B\theta^2).$$

The constants A and B are the same as A and B in equation (1); they can be determined by measuring R_{θ} at the steam point and the sulphur point (444.600°C).

At temperatures below the boiling-point of oxygen the resistance of platinum changes rather slowly with temperature. The resistance of lead changes more rapidly, and resistance thermometers of lead wire have been used.

Thermocouples

Between 630°C and the gold point (1063.0°C) the international temperature scale is expressed in terms of the electromotive force of a thermocouple. The wires of the thermocouple are platinum, and platinum-rhodium alloy (90 per cent Pt.: 10 per cent Rh.). Since the e.m.f. is to be measured on a potentiometer, care must be taken that thermal e.m.f.'s are not set up at the junctions of the thermocouple wires and the copper leads to the potentiometer. To do this three junctions are made, as shown in Fig. 14.3 (a). The junctions of the copper leads to the thermocouple wires are both placed in melting ice. The electro-

motive force of the whole system is then equal to the e.m.f. of two platinum/platinum-rhodium junctions, one in ice and the other at the unknown temperature (Fig. 14.3 (b)).

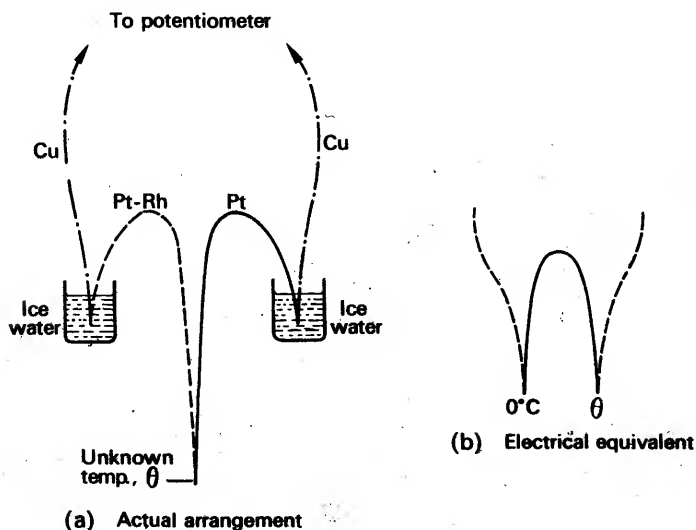


FIG. 14.3. Use of thermocouples.

The international temperature θ corresponding to an e.m.f. E is given by

$$E = a + b\theta + c\theta^2,$$

where a , b and c are constants. The values of the constants are determined by measurements at the gold point (1063.0°C), the silver point (960.8°C), and the temperature of freezing antimony (about 630.3°C). Since the freezing-point of antimony is not a fixed point on the international scale, its value in a given experiment is directly measured with a resistance thermometer. This temperature therefore serves to link the resistance and thermo-electric regions of the temperature scale.

Other Thermocouples

Because of their convenience, thermocouples are used to measure temperatures outside their range on the international scale, when the highest accuracy is not required. The arrangement of three junctions and potentiometer may be used, but for less accurate work the potentiometer may be replaced by a galvanometer G , in the simpler arrangement of Fig. 14.4 (a). The galvanometer scale may be calibrated to read directly in temperatures, the known melting-points of metals like tin and lead being used as subsidiary fixed points. For rough work, particularly at high temperatures, the cold junction may be omitted (Fig. 14.4 (b)). An uncertainty of a few degrees in a thousand is often of no importance.

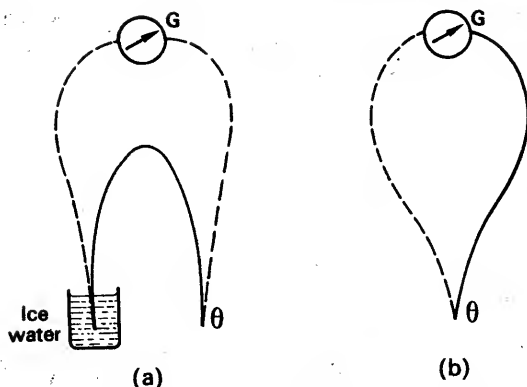


FIG. 14.4. Simple thermojunction thermometer.

E.M.F.'S OF THERMOJUNCTIONS
(In millivolts, cold junction at 0°C)

Temp. of hot junction $^\circ\text{C}$	Pt./Pt.-10% Rh.	Chromel ¹ /Alumel ²	Copper/Constantan ³	Iron/Constantan
100	0.64	4.1	4	5
200	1.44	8.1	9	11
300	2.32	12.2	15	16
400	3.25	16.4	21 ⁵	22
500	4.22	20.6		27
600	5.22	24.9		33
700	6.26	29.1		39
800	7.33	33.3		45
900	8.43	37.4		
1000	9.57	41.3		
1200	11.92 ⁴	48.8		
1400	14.31	55.8		
1600	16.67			

1. Cr-Ni alloy.
2. Al-Ni alloy.
3. 60 per cent Cu, 40 per cent Ni (sometimes called Eureka).
4. Liable to a change of fundamental interval if heated above $1,100^\circ\text{C}$.
5. Cu starts to oxidize.

Intermediate Metals

Fig. 14.5 (a) shows two metals of a thermocouple, A, B, separated by a third metal, C. The metal C may be, for example, a film of the solder used to join A and B. At a given temperature θ , the e.m.f. E of the couple ACB is found, by measurement, to be equal to that of a simple couple AB, formed by twisting or welding the wires together (Fig. 14.5 (b)). This is true provided that the junctions of A to C, and C to B, are both at the temperature θ . An intermediate metal, at a uniform temperature, does not therefore affect the e.m.f. of a thermojunction.

Expanding-liquid Thermometers: Mercury-in-glass

Mercury freezes at -39°C and boils, under atmospheric pressure, at 357°C . Mercury thermometers can be made to read up to about 550°C ,

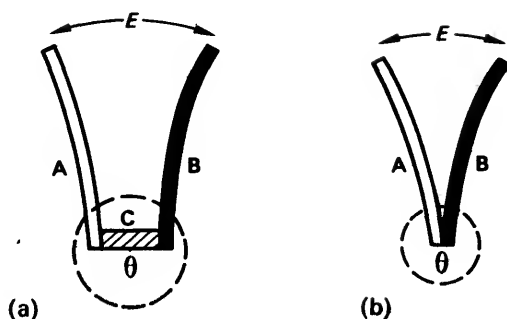


FIG. 14.5. Intermediate metal in a thermojunction.

however, by filling the space above the liquid with nitrogen, which is compressed as the mercury expands, and raises its boiling-point.

A mercury thermometer has a much greater heat capacity than a thermocouple, and cannot follow a rapidly changing temperature. Also glass, when it has been warmed and then cooled, does not immediately contract back to its original volume at the lower temperature. If a low temperature is measured immediately after a high one, the value given by a mercury thermometer tends to be too low. With modern hard glass (Jena glass) this effect is small; but with a cheap thermometer the ice point may be as much as half a degree low if taken immediately after the steam point has been checked.

A mercury thermometer is filled by warming and dipping, as a weight thermometer (p. 283). The mercury in its bulb is then boiled, to drive out air. It is then allowed to cool and draw in the requisite amount of mercury. Finally it is warmed a little above the highest temperature it is to measure, sealed off, and left for about a year—to age. During the ageing period the glass slowly contracts after its strong heating, and at the end of the period the thermometer is calibrated.

Clinical Thermometers

A clinical thermometer has a fine stem, divided into fifths or tenths of a degree, and calibrated over only a small range: 95° – 110° F or 35° – 45° C (Fig. 14.6). The stem is thickened on the side remote from the gradua-

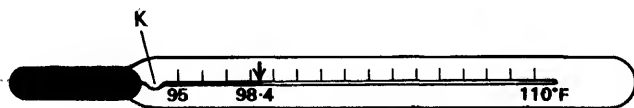


FIG. 14.6. Clinical thermometer.

tions so that it acts as a lens, to magnify the fine mercury thread. At room temperature the mercury retreats right into the bulb. Between the bulb and the graduations is a fine kink K. When the bulb is warmed the mercury is forced through the kink into the stem. But when the bulb is cooled, the mercury does not flow back past the kink—it stays in the stem, so that the temperature can be read at leisure. It is then shaken back into the bulb.

Mercury-in-steel Thermometers

For industrial purposes mercury-in-steel thermometers are used. They consist of a steel bulb **B**, connected by a long steel capillary **S** to a coiled steel tube **C** (Fig. 14.7). The whole is filled with mercury, and when

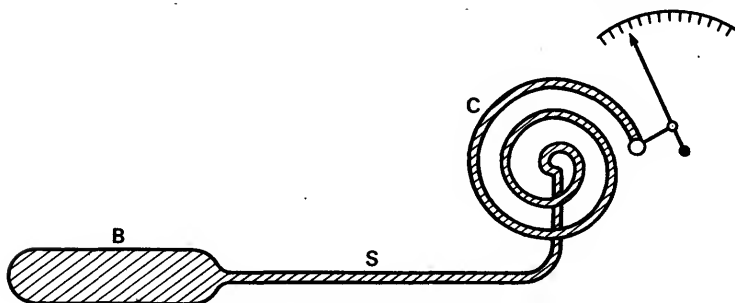


FIG. 14.7. Mercury-in-steel thermometer (*diagrammatic*).

the bulb is warmed the expansion of the mercury makes the coil unwind. The unwinding of the coil actuates a pointer, and indicates the temperature of the bulb. The distance between the bulb and the indicating dial may be many feet.

Alcohol Thermometers

Ethyl alcohol boils at 78°C , and freezes at -115°C . Alcohol thermometers are therefore used in polar regions. Alcohol is also used in some of the maximum and minimum thermometers which we are about to describe.

Maximum and Minimum Thermometers

Meteorologists observe the highest and lowest temperatures reached by the air day and night. They use a *maximum thermometer* which is a mercury thermometer containing a small glass index **I** (Fig. 14.8 (a)). The thermometer is laid horizontally, in a louvred screen. When the temperature rises the mercury pushes the index along, but when the temperature falls the mercury leaves the index behind. The maximum temperature is therefore shown by the end of the index nearer the mercury. After each observation, the index is brought back to the mercury by tilting the thermometer.

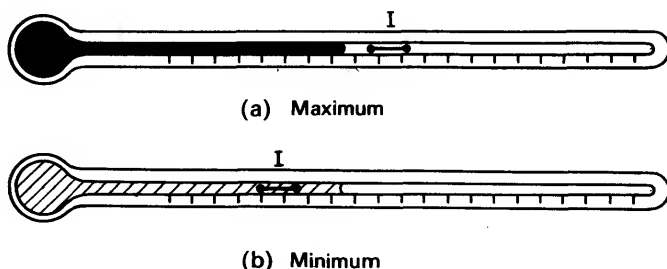


FIG. 14.8. Meteorological thermometers.

A *minimum thermometer* is similar to a maximum, except that it contains alcohol instead of mercury. When the alcohol expands it flows past the index, but when it contracts it drags the index back, because of its surface tension (Fig. 14.8 (b)). The end of the index nearer the meniscus therefore shows the minimum temperature. The index is re-set by tilting.

A combined *maximum and minimum thermometer* was invented by Six in 1782. Its construction is shown in Fig. 14.9. The bulb A, and

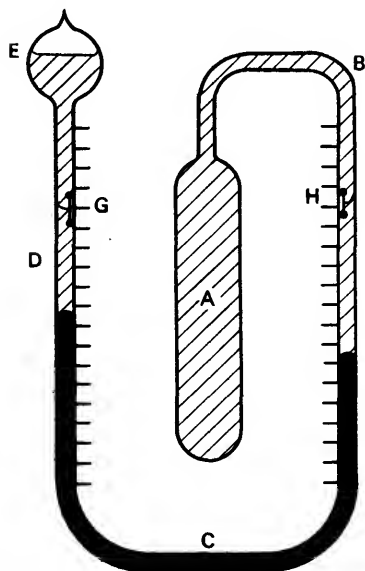


FIG. 14.9. Combined maximum and minimum thermometer.

the part B of the stem, contain alcohol; so does the part D of the stem, and the lower part of the bulb E. The part C of the stem contains mercury, and the upper part of the bulb E contains air and saturated alcohol vapour. The indices G and H are made of iron, and are fitted with springs pressing against the walls of the stem. When the temperature rises, the expansion of the large volume of alcohol in A forces the mercury round, and compresses the gases in E. The mercury pushes the index G up the tube. When the temperature falls, the alcohol in A contracts, and the mercury retreats. But the spring holds the index G, and the alcohol flows past it. Thus the bottom of G shows the maximum temperature. Similarly, the bottom of H shows the minimum. The indices are re-set

by dragging them along from the outside with a magnet.

Bimetal Strip Thermometers

If a bimetal strip is wound into a spiral, with the more expansible metal on the inside, then the spiral will uncoil as the temperature rises. The movement of the spiral can be made to turn a pointer, and to act as a thermometer. Instruments of this kind do not hold their calibration as well as liquid-in-glass thermometers.

PYROMETERS

High temperatures are usually measured by observing the radiation from the hot body, and the name *Pyrometry* is given to this measurement. Before describing pyrometers, however, we may mention some other, rough, methods which are sometimes used. One method is to insert in the furnace a number of ceramic cones, of slightly different compositions; their melting-points increase from one to the next by about 20°C . The temperature of the furnace lies between the melting-

points of adjacent cones, one of which softens and collapses, and the other of which does not.

The temperature of steel, when it is below red heat, can be judged by its colour, which depends on the thickness of the oxide film upon it. Temperatures below red heat can also be estimated by the use of paints, which change colour at known temperatures.

Radiation pyrometers can be used only above red-heat (about 600°C). They fall into two classes:

- (i) *total radiation pyrometers*, which respond to the total radiation from the hot body, heat and light;
- (ii) *optical pyrometers*, which respond only to the visible light.

Optical Pyrometers

Fig. 14.10 illustrates the principle of the commonest type of optical pyrometer, called a *disappearing filament pyrometer*. It consists essen-

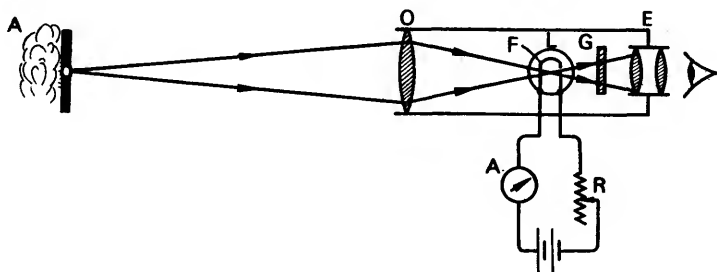


FIG. 14.10. Optical radiation pyrometer (*not to scale*).

tially of a low power telescope, OE, and a tungsten filament lamp L. The eyepiece E is focused upon the filament F. The hot body A whose temperature is to be found is then focused by the lens O so that its image lies in the plane of F. The light from both the filament and the hot body passes through a filter of red glass G before reaching the eye. If the body is brighter than the filament, the filament appears dark on a bright ground. If the filament is brighter than the body, it appears bright on a dark ground. The temperature of the filament is adjusted, by adjusting the current through it, until it merges as nearly as possible into its background. It is then as bright as the body. The rheostat R which adjusts the current is mounted on the body of the pyrometer, and so is the ammeter A which measures the current. The ammeter is calibrated directly in degrees Centigrade or Fahrenheit. A pyrometer of this type can be adjusted to within about 5°C at 1000°C ; more elaborate types can be adjusted more closely.

The range of an optical pyrometer can be extended by introducing a filter of green glass between the objective O and the lamp L; this reduces the brightness of the red light. A second scale on the ammeter is provided for use when the filter is inserted.

The scale of a radiation pyrometer is calibrated by assuming that the radiation is black-body radiation (p. 352). If—as usual—the hot body is not black, then it will be radiating less intensely than a black

body at the same temperature. Conversely, a black body which radiates with the same intensity as the actual body will be cooler than the actual body. Thus the temperature indicated by the pyrometer will be lower than the true temperature of the actual, not black, body. A correction must be applied to the pyrometer reading, which depends on the spectral emissivity of the body for red light. The wavelength λ for which the spectral emissivity e_λ must be known is the average wavelength of light transmitted by the red filter—usually about 6500 Å.U.

The following tables give e_λ for various substances, and the corrections to be added for various values of e_λ .

SPECTRAL EMISSIVITIES, e_λ
(At 650 nm)

Substance	Solid	Liquid	Substance	Solid	Liquid
Carbon	0.85	—	Nickel	0.35	—
Copper	0.1	0.15	Nickel oxidized . .	0.9	—
Copper oxidized . .	0.7	—	Platinum	0.35	0.35
Gold	0.15	0.2	Silver	0.1	—
Iron	0.35	0.35	Slag	—	0.65
Iron oxidized . . .	0.95	—	Tungsten	0.45	—
Nichrome	0.9	—			

OPTICAL PYROMETER CORRECTIONS
($\lambda = 650 \text{ nm}$)

[To be added to observed temperature]

Observed temp., °C							
	0.3	0.4	0.5	0.6	0.7	0.8	0.9
600	44	34	26	18	13	8	4
800	67	50	37	27	19	12	6
1000	95	71	53	39	27	17	8
1200	129	96	71	52	36	22	10
1400	169	125	93	67	46	28	13
1600	214	159	117	85	58	35	17
1800	265	196	145	105	72	44	20
2000	322	238	176	127	87	53	25
2500	495	362	266	190	131	78	38
3000	713	516	377	269	183	110	53

Total Radiation Pyrometers

Total radiation pyrometers are less common than optical pyrometers. As we shall see, they can only be used when the source of radiation is of considerable size—such as the open door of a furnace—whereas an optical pyrometer can be used on a very small body such as a lamp-filament.

Fig. 14.11 illustrates the principle of a Féry total radiation pyrometer. The blackened tube A is open at the end B; at the other end C it carries an eye-piece E. D is a thermocouple attached to a small blackened

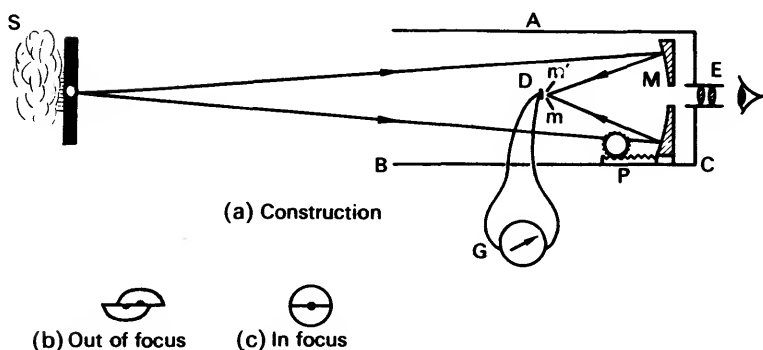


FIG. 14.11. Fény total radiation pyrometer.

disc of copper, which faces the end C of the tube and is shielded from direct radiation. M is a gold-plated mirror, pierced at the centre to allow light to reach the eye-piece, and moveable by a rack and pinion P.

In use, the eyepiece is first focused upon the disc D. The mirror M is then adjusted until the furnace is also focused upon D. Since a body which is black or nearly so shows no detail, focusing it upon D by simply looking at the image would be almost impossible. To make the focusing easier, two small plane mirrors m' , m are fitted in front of D. They are inclined with their normals at about 5° to the axis of the tube, and are pierced with semi-circular holes to allow radiation from M to reach the disc. The diameter of the resulting circular hole is less than that of the disc. When the source of heat is not focused on the disc, the two mirrors appear as at (b) in the figure; when the focusing is correct, they appear as at (c). The source must be of such a size that its image completely fills the hole.

The radiation from the source warms the junction and sets up an electromotive force. A galvanometer G connected to the junction is then deflected, and can be calibrated to read directly the temperature of the source.

The calibration gives the correct temperature if the source is a black body. If the source is not black, its total radiation is equal to that of a black body at some lower temperature; the pyrometer therefore reads too low. If the total emissivity of the source is known, a correction can be made for it. This correction is greater than it would be if an optical pyrometer were used, that is to say, departure from perfect blackness causes less error in an optical pyrometer than in a total radiation one.

The Foster Pyrometer

Another type of total radiation pyrometer is the Foster fixed-focus instrument (Fig. 14.12); it also uses a thermojunction with blackened disc. A is an open diaphragm, so placed that it and the thermojunction D are at conjugate foci of the mirror M. The thermojunction then collects all the radiation entering through A but, since it is much smaller than A, it is raised to a higher temperature than if its disc were the size of A. The radiation entering through A is limited to that within

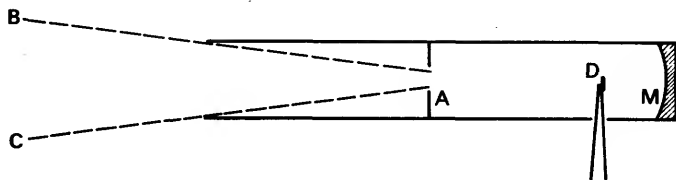


FIG. 14.12. Fixed focus total radiation pyrometer.

the cone ABC; as long as the whole of this cone is intersected by the hot source, the amount of radiation reaching the thermocouple is independent of the distance to the source (compare the use of a cone with a thermopile, p. 350).

Comparison of Pyrometers: the International Scale

Total radiation and optical pyrometers agree within the limits of experimental error— $\frac{1}{2}^{\circ}$ at 1750°C , about 4° at 2800°C . The choice between them is decided solely by convenience. The international temperature scale above the gold-point (1063.0°C) is defined in terms of an optical pyrometer.

Extension of Range by Sectored Disc

The range of a radiation pyrometer can be extended by cutting down the radiation admitted to it. A disc from which an angle θ

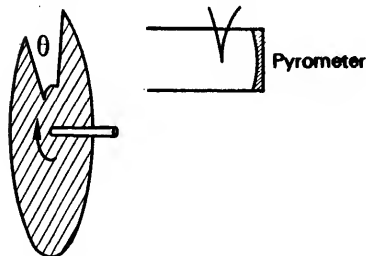


FIG. 14.13. Sectored disc with radiation pyrometer.

radians has been cut out is rotated in front of the pyrometer, as shown in Fig. 14.13, so that the radiation entering is cut down in the ratio $\theta/2\pi$. The pyrometer then indicates a temperature T_1 , which is less than the true temperature T_2 of the source. The temperatures T are expressed in K to simplify the calculation which follows.

If the pyrometer is of the total radiation type, then we can use Stefan's law. The radiation from a body at T_2 K is proportional to T_2^4 . The pyrometer receives radiation represented by the temperature T_1 , and therefore proportional to T_1^4 .

Therefore
$$\frac{\theta}{2\pi} = \frac{T_1^4}{T_2^4}$$

whence
$$T_2 = T_1 \left(\frac{2\pi}{\theta} \right)^{\frac{1}{4}}.$$

In this way the surface temperature of the sun has been estimated. The value found agrees with that estimated from the wavelength of the sun's most intense radiation (p. 353); it is about 6000 K.

A sectored disc can be used to extend the range of an optical pyrometer, but the calculation is more difficult than for a total radiation pyrometer.

EXAMPLES

1. How is Celsius temperature defined (a) on the scale of a constant-pressure gas thermometer, (b) on the scale of a platinum resistance thermometer? A constant mass of gas maintained at constant pressure has a volume of 200.0 cm^3 at the temperature of melting ice, 273.2 cm^3 at the temperature of water boiling under standard pressure, and 525.1 cm^3 at the normal boiling-point of sulphur. A platinum wire has resistances of 2.000 , 2.778 and 5.280 ohms at the same temperatures. Calculate the values of the boiling-point of sulphur given by the two sets of observations, and comment on the results. (N.)

First part (a). The temperature θ on the gas thermometer scale is given by

$$\theta = \frac{V_\theta - V_0}{V_{100} - V_0} \times 100,$$

where V_θ , V_0 , V_{100} are the respective volumes of the gas at constant pressure at the temperature concerned, the temperature of melting ice, and the temperature of steam at 76 cm mercury pressure. (b) The temperature θ_p on the platinum resistance thermometer scale is given by

$$\theta_p = \frac{R_\theta - R_0}{R_{100} - R_0} \times 100,$$

where R_θ , R_0 , R_{100} are the respective resistances of the platinum at the temperature concerned, the temperature of melting ice, and the temperature of steam at 76 cm mercury.

Second part. On the gas thermometer scale, the boiling-point of sulphur is given by

$$\begin{aligned} \theta &= \frac{525.1 - 200.0}{273.2 - 200.0} \times 100 \\ &= 444.1^\circ\text{C}. \end{aligned}$$

On the platinum resistance thermometer scale, the boiling-point is given by

$$\begin{aligned} \theta_p &= \frac{5.280 - 2.000}{2.778 - 2.000} \times 100 \\ &= 421.6^\circ\text{C}. \end{aligned}$$

The temperatures recorded on the thermometers are therefore different. This is due to the fact that the variation of gas pressure with temperature at constant volume is different from the variation of the electrical resistance of platinum with temperature.

2. Explain how a Celsius temperature scale is defined, illustrating your answer by reference to a platinum resistance thermometer.

The resistance R_t of a platinum wire at temperature $t^\circ\text{C}$, measured on the gas scale, is given by $R_t = R_0(1 + at + bt^2)$, where $a = 3.800 \times 10^{-3}$ and $b = -5.6 \times 10^{-7}$. What temperature will the platinum thermometer indicate when the temperature on the gas scale is 200°C ? (O. & C.)

First part. The temperature θ_p in $^\circ\text{C}$ on a resistance thermometer scale is given by the relation

$$\theta_p = \frac{R_\theta - R_0}{R_{100} - R_0} \times 100,$$

where R_θ , R_0 , R_{100} are the respective resistances at the temperature concerned, at 0°C , and at 100°C .

Second part. $R_t = R_0(1 + at + bt^2)$

$$\therefore R_{200} = R_0(1 + 200a + 200^2b)$$

and

$$R_{100} = (R_0(1 + 100a + 100^2b)).$$

$$\begin{aligned}\therefore \theta_p &= \frac{R_{200} - R_0}{R_{100} - R_0} \times 100 \\ &= \frac{R_0(1 + 200a + 200^2b) - R_0}{R_0(1 + 100a + 100^2b) - R_0} \times 100 \\ &= \frac{200a + 200^2b}{a + 100b} = \frac{200(a + 200b)}{a + 100b} \\ &= 200 \frac{(3.8 \times 10^{-3} - 11.2 \times 10^{-5})}{3.8 \times 10^{-3} - 5.6 \times 10^{-5}} \\ &= \frac{200 \times 0.003688}{0.003744} = 197^\circ\text{C}.\end{aligned}$$

EXERCISES 14

1. How is a scale of temperature defined? What is meant by a temperature of 15°C ?

On what evidence do you accept the statement that there is an absolute zero of temperature at about -273°C ?

In a special type of thermometer a fixed mass of gas has a volume of 100.0 cm^3 and a pressure of 81.6 cm of mercury at the ice point, and volume 124.0 units, with pressure 90.0 units at the steam point. What is the temperature when its volume is 120.0 units and pressure 85.0 units, and what value does the scale of this thermometer give for absolute zero? Explain the principle of your calculation. (*O. & C.*)

2. Describe the structure of (a) a platinum resistance thermometer, (b) an optical pyrometer. Explain, giving details of the auxiliary electrical circuits required, how you would use each type of thermometer to measure a temperature, assuming that the instrument has already been calibrated over the required range. (*L.*)

3. Explain what is meant by a change in temperature of 1 deg C on the scale of a platinum resistance thermometer.

Draw and label a diagram of a platinum resistance thermometer together with a circuit in which it is used.

Give two advantages of this thermometer and explain why, in its normal form, it is unsuited for measurement of varying temperatures.

The resistance R_t of platinum varies with the temperature $t^\circ\text{C}$ as measured by a constant volume gas thermometer according to the equation

$$R_t = R_0(1 + 8,000\alpha t - \alpha^2 t^2)$$

where α is a constant. Calculate the temperature on the platinum scale corresponding to 400°C on this gas scale. (*N.*)

4. Give the essential steps involved in setting up a scale of temperature. Explain why scales based on different properties do not necessarily agree at all temperatures. At what temperature or temperatures do these different scales agree?

The volume of some air at constant pressure, and also the length of an iron rod, are measured at 0°C and again at 100°C with the following results:

	0°C	100°C
Volume of air (cm ³)	28.5	38.9
Length of rod (cm)	100.00	100.20

Calculate (a) the absolute zero of this air thermometer scale, and (b) the length of the iron rod at this temperature if its expansion is uniform according to the air scale. (C.)

5. Describe the structure of a constant volume gas thermometer. Describe also the method of calibrating it and using it to determine the boiling point of a salt solution.

Compare this thermometer as a means of measuring temperature with (a) a mercury-in-glass thermometer, (b) a thermoelectric thermometer. (L.)

6. State what is meant by a temperature on the centigrade (Celsius) scale of a platinum resistance thermometer.

Point out the relative merits of (a) a platinum resistance thermometer, and (b) a thermoelectric thermometer for measuring (i) the rise in temperature of the water flowing through a continuous flow calorimeter, and (ii) the temperature of a small crystal as it is being heated rapidly. (N.)

7. Three types of thermometer in common use are based on (a) the expansion of a fluid, (b) the production of an electromotive force, (c) the variation of electrical resistance. Describe briefly one example of each and the way in which it is used. In each case, state how a value of the temperature on a centigrade scale is deduced from the quantities actually measured.

If all three of the thermometers you have described were used to measure the temperature of the same object, would they give the same result? Give reasons for your answer. (O. & C.)

8. Explain the principle underlying the establishment of a centigrade temperature scale in terms of some suitable physical property.

What type of thermometer would you choose for use in experiments involving (a) the plotting of a cooling curve for naphthalene in the region of its melting point, (b) finding the boiling point of oxygen, (c) the measurement of the thermal conductivity of a small crystal? In each instance give reasons for your choice.

If the resistance R_t of the element of a resistance thermometer at a temperature of $t^\circ\text{C}$ on the ideal gas scale is given by $R_t = R_0(1 + At + Bt^2)$, where R_0 is the resistance at 0°C and A and B are constants such that $A = -6.50 \times 10^{-3}B$, what will be the temperature on the scale of the resistance thermometer when $t = 50.0^\circ\text{C}$? (L.)

9. Explain briefly what is meant by a temperature of $\theta^\circ\text{C}$ on (a) the mercury-in-glass scale, (b) the constant-pressure hydrogen scale, (c) the platinum resistance scale.

For what range of temperatures could (i) a platinum resistance thermometer, (ii) an optical pyrometer be employed? Describe the structure and the method of use of one of these instruments. (L.)

10. Explain the principle of a constant volume gas thermometer and describe a simple instrument suitable for measurements in the range 0° to 100°C . What factors determine (a) the sensitivity and (b) the accuracy of the instrument you describe?

A certain gas thermometer has a bulb of volume 50 cm^3 connected by a capillary tube of negligible volume to a pressure gauge of volume 5.0 cm^3 . When the bulb is immersed in a mixture of ice and water at 0°C , with the pressure gauge at room temperature (17°C), the gas pressure is 700 mm Hg. What will be the pressure

when the bulb is raised to a temperature of 50°C if the gauge is maintained at room temperature? You may assume that the gas is ideal and that the expansion of the bulb can be neglected.

11. Give a brief account of the principles underlying the establishment of a scale of temperature and explain precisely what is meant by the statements that the temperature of a certain body is (a) $t^{\circ}\text{C}$ on the constant volume air scale, (b) $t_p^{\circ}\text{C}$ on the platinum resistance scale, and (c) $t_T^{\circ}\text{C}$ on the Cu-Fe thermocouple scale. Why are these three temperatures usually different?

Describe an optical pyrometer and explain how it is used to measure the temperature of a furnace. (N.)

12. What is the general method of calibrating any type of thermometer? Describe *briefly* the method of using *three* of the following temperature measuring devices, and give the temperature range in which they are most usefully employed: (i) platinum resistance thermometer, (ii) mercury-in-glass thermometer, (iii) helium gas thermometer, (iv) optical pyrometer. (L.)

13. Explain the precautions taken in verifying the position of the fixed points on the stem of a mercury-in-glass thermometer.

Describe a constant volume air thermometer. How could such a thermometer be used to determine the melting-point of a solid such as naphthalene? What advantage has the gas thermometer and for what purpose is it used? (L.)

14. Tabulate various physical properties used for measuring temperature. Indicate the temperature range for which each is suitable.

Discuss the fact that the numerical value of a temperature expressed on the scale of the platinum resistance thermometer is not the same as its value on the gas scale except at the fixed points.

If the resistance of a platinum thermometer is 1.500 ohms at 0°C , 2.060 ohms at 100°C and 1.788 ohms at 50°C on the gas scale, what is the difference between the numerical values of the latter temperature on the two scales? (N.)

15. Explain how a centigrade temperature scale is defined, illustrating your answer by reference to a platinum resistance thermometer.

The resistance R_t of a platinum wire at temperature $t^{\circ}\text{C}$, measured on the gas scale, is given by $R_t = R_0(1 + at + bt^2)$, where $a = 4.000 \times 10^{-3}$ and $b = -6.0 \times 10^{-7}$. What temperature will the platinum thermometer indicate when the temperature on the gas scale is 300°C ?

16. Describe how you would use *either* (i) a constant-volume or (ii) a constant-pressure air thermometer to calibrate a mercurial thermometer. If the difference of mercury level in a constant-volume air thermometer is -2 cm when the temperature of the bulb is 10°C and $+22$ cm when the bulb is at 100°C , what is the height of the barometer? (L.)

PART THREE

Optics and Sound

OPTICS

chapter fifteen

Introduction

If you wear spectacles you will appreciate particularly that the science of Light, or *Optics* as it is often called, has benefited people all over the world. The illumination engineer has developed the branch of Light dealing with light energy, and has shown how to obtain suitable lighting conditions in the home and the factory, which is an important factor in maintaining our health. Microscopes, used by medical research workers in their fight against disease; telescopes, used by seamen and astronomers; and a variety of optical instruments which incorporate lenses, mirrors, or glass prisms, such as cameras, driving mirrors, and binoculars, all testify to the scientist's service to the community.

In mentioning the technical achievements in Light, however, it must not be forgotten that the science of Light evolved gradually over the past centuries; and that the technical advances were developed from the experiments and theory on the *fundamental* principles of the subject made by scientists such as NEWTON, HUYGENS, and FRESNEL.

Light Travels in Straight Lines. Eclipses and Shadows

When sunlight is streaming through an open window into a room, observation shows that the edges of the beam of light, where the shadow begins, are straight. This suggests that *light travels in straight lines*, and on this assumption the sharpness of shadows and the formation of eclipses can be explained. Fig. 15.1 (i) illustrates the eclipse of the sun, S,

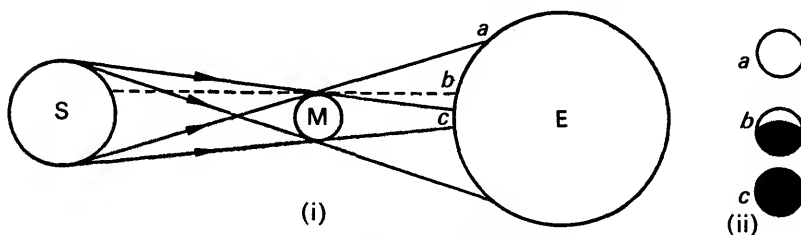


FIG. 15.1. Eclipse of Sun (*not to scale*).

our natural source of light, when the moon, M, passes between the sun and the earth, E. The moon is a non-luminous object which does not allow light to pass through it, and hence the boundaries of the shadows formed by M on the earth are obtained by drawing lines from S which

touch the edge of M. Consequently there is a total eclipse of the sun at *c* on the earth, a partial eclipse at *b*, and no eclipse at *a*. Fig. 15.1 (ii) illustrates the appearance of the sun in each case.

Light Rays and Beams

Light is a form of energy. We know this is the case because plants and vegetables grow when they absorb sunlight. Further, electrons are ejected from certain metals when light is incident on them, showing that there was some energy in the light; this phenomenon is the basis of the *photoelectric cell* (p. 1077). Substances like wood or brick which allow no light to pass through them are called "opaque" substances; unless an opaque object is perfectly black, some of the light falling on it is reflected (p. 391). A "transparent" substance, like glass, is one which allows some of the light energy incident on it to pass through, the remainder of the energy being absorbed and (or) reflected.

A ray of light is the direction along which the light energy travels; and though rays are represented in diagrams by straight lines, in practice a ray has a finite width. A **beam** of light is a collection of rays. A search-light emits a *parallel beam* of light, and the rays from a point on a very distant object like the sun are substantially parallel, Fig. 15.2 (i). A lamp

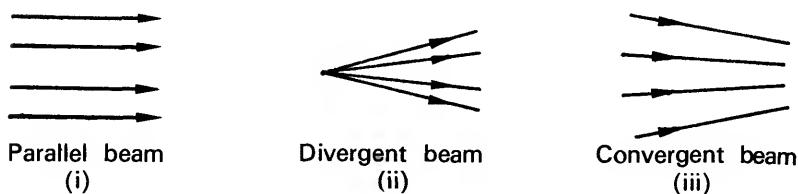


FIG. 15.2. Beams of light.

emits a *divergent beam* of light; while a source of light behind a lens, as in a projection lantern, can provide a *convergent beam*, Fig. 15.2 (ii), (iii).

The Velocity of Light

The velocity of light is constant for a given medium, such as air, water, or glass, and has its greatest magnitude, about 3.0×10^8 metres per second, in a vacuum. The velocity of light in air differs only slightly from its velocity in a vacuum, so that the velocity in air is also about 3.0×10^8 metres per second. The velocity of light in glass is about 2.0×10^8 metres per second; in water it is about 2.3×10^8 metres per second. On account of the difference in velocity in air and glass, light changes its direction on entering glass from air (see *Refraction*, p. 679). Experiments to determine the velocity of light are discussed later, p. 551.

The Human Eye

When an object is seen, light energy passes from the object to the observer's eyes and sets up the sensation of vision. The eye is thus sensitive to light (or luminous) energy. The eye contains a *crystalline lens*, L, made of a gelatinous transparent substance, which normally throws an image of the object viewed on to a sensitive "screen" R at the back of the eye-ball, called the *retina*, Fig. 15.3. Nerves on the retina

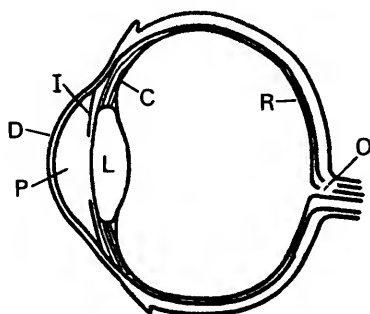


FIG. 15.3. The eyeball.

are joined to the *optic nerve*, O, which carries the sensation produced by the image to the brain. The *iris*, I, is a diaphragm with a circular hole in the middle called the *pupil*, P, which contracts when the light received by the eye is excessive and painful to the eye. The colour of a person's eyes is the colour of the iris; the pupil is always black because no light returns from the interior of the eye-ball. A weak salt solution, called the *aqueous humour*, is present on the left of the lens L, and between L and the retina is a gelatinous substance called the *vitreous humour*. The transparent spherical bulge D in front of L is made of tough material, and is known as the *cornea*.

The *ciliary muscles*, C, enable the eye to see clearly objects at different distances, a property of the eye known as its "power of accommodation". The ciliary muscles are attached to the lens L, and when they contract, the lens' surfaces are pulled out so that they bulge more; in this way a near object can be focused clearly on the retina and thus seen distinctly. When a very distant object is observed the ciliary muscles are relaxed, and the lens' surfaces are flattened.

The use of two eyes gives a three-dimensional aspect of the object or scene observed, as two slightly different views are imposed on the retinae; this gives a sense of distance not enjoyed by a one-eyed person.

Direction of Image seen by Eye

When a fish is observed in water, rays of light coming from a point^c such as O on it pass from water into air, Fig. 15.4 (i). At the boundary of

the water and air, the rays OA, OC proceed along new directions AB, CD respectively and enter the eye. Similarly, a ray OC from an object O observed in a mirror is reflected along a new direction CD and enters the eye, Fig. 15.4 (ii). These phenomena are studied more fully later, but

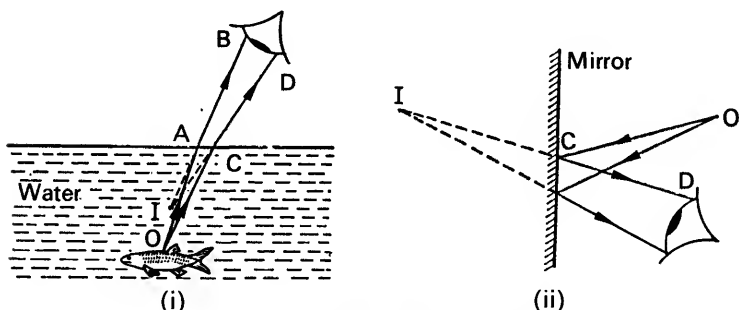


FIG. 15.4. Images observed by eye.

the reader should take careful note that the eye sees an object *in the direction in which the rays enter the eye*. In Fig. 15.4 (i), for example, the object O is seen in the water at I, which lies on BA and DC produced slightly on the right of O; in Fig. 15.4 (ii), is seen behind the mirror at I, which lies on DC produced. In either case, all rays from O which enter the eye appear to come from I, which is called the image of O.

Reversibility of Light

If a ray of light is directed along DC towards a mirror, experiment shows that the ray is reflected along the path CO, Fig. 15.4 (ii). If the ray is incident along OC, it is reflected along CD, as shown. Thus if a light ray is reversed it always travels along its original path, and this is known as *the principle of the reversibility of light*. In Fig. 15.4 (i), a ray BA in air is refracted into the water along the path AO, since it follows the reverse path to OAB. We shall have occasion to use the principle of the reversibility of light later in the book.

chapter sixteen

Reflection at plane surfaces

HIGHLY-POLISHED metal surfaces reflect about 80 to 90 per cent of the light incident on them; *mirrors* in everyday use are therefore usually made by depositing silver on the back of glass. In special cases the front of the glass is coated with the metal; for example, the largest reflector in the world is a curved mirror nearly 5 metres across, the front of which is coated with aluminium (p. 544). Glass by itself will also reflect light, but the percentage reflected is small compared with the case of a silvered surface; it is about 5 per cent for an air-glass surface.

Laws of Reflection

If a ray of light, AO , is incident on a plane mirror XY at O , the angle AON made with the *normal* ON to the mirror is called the “angle of incidence”, i , Fig. 16.1. The angle BON made by the reflected ray OB with the normal is called the “angle of reflection”, r ; and experiments with a ray-box and a plane mirror, for example, show that:

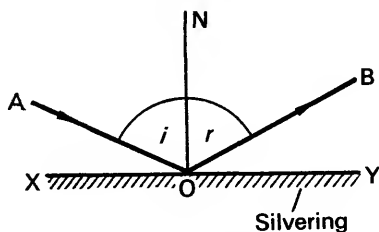


FIG. 16.1. Plane mirror.

(1) *The reflected ray, the incident ray, and the normal to the mirror at the point of incidence all lie in the same plane.*

(2) *The angle of incidence = the angle of reflection.*

These are called the two *laws of reflection*, and they were known to PLATO, who lived about 400 B.C.

Regular and Diffuse Reflection

In the case of a plane mirror or glass surface, it follows from the laws of reflection that a ray incident at a given angle on the surface is reflected in a definite direction. Thus a parallel beam of light incident on a plane mirror in the direction AO is reflected as a parallel beam in the direction OB , and this is known as a case of *regular reflection*, Fig. 16.2 (i). On the other hand, if a parallel beam of light is incident on a sheet of paper in a direction AO , the light is reflected in all different directions from the paper: this is an example of *diffuse reflection*, Fig. 16.2 (ii). Objects in everyday life, such as flowers, books, people, are seen by light diffusely reflected from them. The explanation of the diffusion of light is that the

surface of paper, for example, is not perfectly smooth like a mirrored surface; the "roughness" in a paper surface can be seen with a micro-

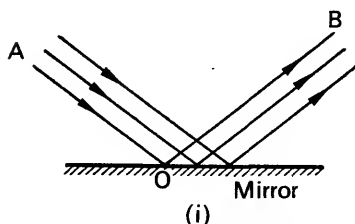


FIG. 16.2 (i). Regular reflection.

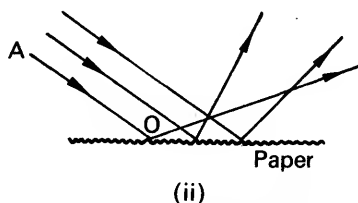


FIG. 16.2 (ii). Diffuse reflection.

scope. At each point on the paper the laws of reflection are obeyed, but the angle of incidence varies, unlike the case of a mirror.

Deviation of Light at Plane Mirror Surface

Besides other purposes, plane mirrors are used in the sextant (p. 395), in simple periscopes, and in signalling at sea. These instruments utilise the property of a plane mirror to deviate light from one direction to another.

Consider a ray AO incident at O on a plane mirror XY , Fig. 16.3 (i). The angle AOX made by AO with XY is known as the *glancing angle*, g , with the mirror; and since the angle of reflection is equal to the angle of incidence, the glancing angle BOY made by the reflected ray OB with the mirror is also equal to g .

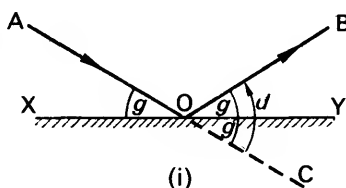


FIG. 16.3 (i).

Deviation of light at a plane mirror.

The light has been deviated from a direction AO to a direction OB . Thus if OC is the extension of AO , the *angle of deviation*, d , is angle COB . Since angle $COY =$ vertically opposite angle XOA , it follows that

$$d = 2g \quad (1);$$

so that, in general, *the angle of deviation of a ray by a plane surface is twice the glancing angle.*

Deviation of Reflected Ray by Rotated Mirror

Consider a ray AO incident at O on a plane mirror M_1 , a being the glancing angle with M_1 , Fig. 16.3 (ii). If OB is the reflected ray, then, as shown above, the angle of deviation $COB = 2g = 2a$.

Suppose the mirror is rotated through an angle θ to a position M_2 , the direction of the incident ray AO being *constant*. The ray is now reflected from M_2 , in a direction OP , and the glancing angle with M_2 is $(\alpha + \theta)$. Hence the new angle of deviation $\angle COP = 2g = 2(\alpha + \theta)$. The reflected ray has thus been rotated through an angle $\angle BOP$ when the mirror rotated through an angle θ ; and since

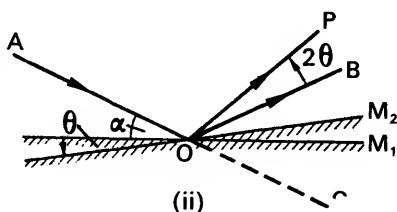


FIG. 16.3 (ii). Rotation of reflected ray.

$$\begin{aligned}\angle BOP &= \angle COP - \angle COB, \\ \text{then } \angle BOP &= 2(\alpha + \theta) - 2\alpha = 2\theta.\end{aligned}$$

Thus, if the direction of an incident ray is constant, the angle of rotation of the reflected ray is twice the angle of rotation of the mirror. If the mirror rotates through 4° , the reflected ray turns through 8° , the direction of the incident ray being kept unaltered.

Optical Lever in Mirror Galvanometer

In a number of instruments a beam of light is used as a "pointer", which thus has a negligible weight and is sensitive to deflections of the moving system inside the instrument. In a mirror galvanometer, used for measuring very small electric currents, a small mirror M_1 is rigidly attached to a system which rotates when a current flows in it, and a beam of light from a fixed lamp L shines on the mirror, Fig. 16.4. If the light is incident normally on the mirror at A , the beam is reflected directly back, and a spot of light is obtained at O on a graduated scale S

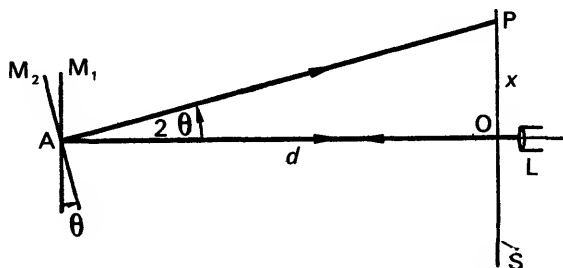


FIG. 16.4. Optical lever principle.

placed just above L . Suppose that the moving system, to which the mirror is attached, undergoes a rotation θ . The mirror is then rotated through this angle to a position M_2 , and the spot of light is deflected through a distance x , say to a position P on the scale.

Since the direction OA of the incident light is constant, the rotation of the reflected ray is twice the angle of rotation of the mirror (p. 393). Thus angle $OAP = 2\theta$. Now $\tan 2\theta = x/d$, where d is the distance OA . Thus 2θ can be calculated from a knowledge of x and d , and hence θ

is obtained. If 2θ is small, then $\tan 2\theta$ is approximately equal to 2θ in radians, and in this case θ is equal to $x/2d$ radians.

In conjunction with a mirror, a beam of light used as a "pointer" is known as an "optical lever". Besides a negligible weight, it has the advantage of magnifying by two the rotation of the system to which the mirror is attached, as the angle of rotation of the reflected light is twice the angle of rotation of the mirror. An optical lever can be used for measuring small increases of length due to the expansion or contraction of a solid.

Deviation by Successive Reflections at Two Inclined Mirrors

Before we can deal with the principle of the sextant, the deviation of light by successive reflection at two inclined mirrors must be discussed.

Consider two mirrors, XO, XB, inclined at an angle θ , and suppose AO is a ray incident on the mirror XO at a glancing angle α , Fig. 16.5 (i). The reflected ray OB then also makes a glancing angle α with OX, and from our result on p. 392, the angle of deviation produced by XO in a clockwise direction (angle LOB) = 2α .

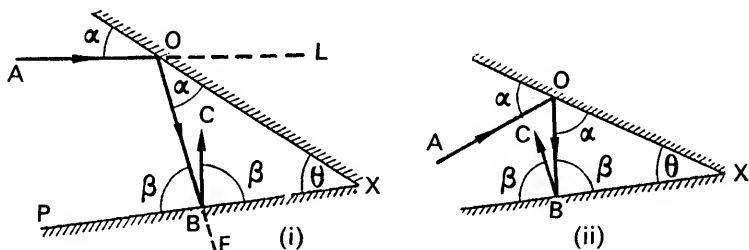


FIG. 16.5. Successive reflection at two plane mirrors.

Suppose OB is incident at a glancing angle β on the second mirror XB. Then, if the reflected ray is BC, the angle of deviation produced by this mirror (angle EBC) = 2β , in an anti-clockwise direction. Thus the net deviation D of the incident ray AO produced by both mirrors = $2\beta - 2\alpha$, in an anti-clockwise direction.

Now from triangle OBX,

$$\text{angle PBO} = \text{angle BOX} + \text{angle BXO},$$

$$\text{i.e.,} \quad \beta = \alpha + \theta$$

Thus $\theta = \beta - \alpha$, and hence

$$D = 2\beta - 2\alpha = 2\theta.$$

But θ is a constant when the two mirrors are inclined at a given angle. Thus, no matter what angle the incident ray makes with the first mirror, the deviation D after two successive reflections is constant and equal to twice the angle between the mirrors.

Fig. 16.5 (ii) illustrates the case when the ray BC reflected at the second mirror travels in an opposite direction to the incident ray AO, unlike the case in Fig. 16.5 (i). In Fig. 16.5 (ii), the net deviation, D , after two successive reflections in a clockwise direction is $2\alpha + 2\beta$. But $\alpha + \beta = 180^\circ - \theta$. Hence $D = 2\alpha + 2\beta = 360^\circ - 2\theta$. Thus the deviation, D , in an anti-clockwise direction is 2θ , the same result as obtained above.

Principle of the Sextant

The *sextant* is an instrument used in navigation for measuring the angle of elevation of the sun or stars. It consists essentially of a fixed glass B, silvered on a vertical half, and a silvered mirror O which can be rotated about a horizontal axis. A small fixed telescope T is directed towards B, Fig. 16.6.

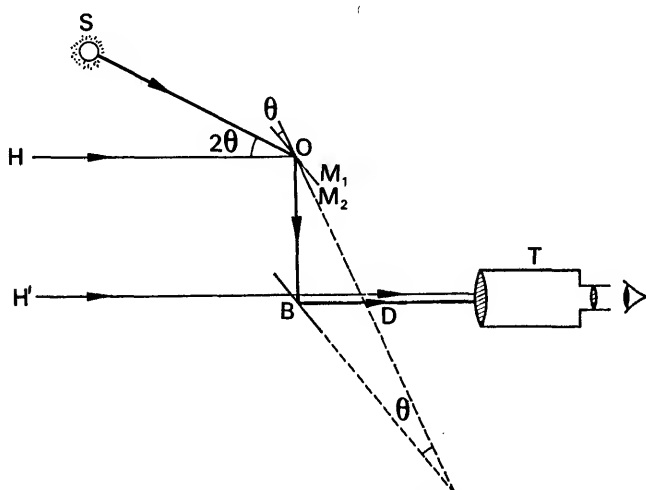


FIG. 16.6. Sextant principle.

Suppose that the angle of elevation of the sun, S, is required. Looking through T, the mirror O is turned until the view H' of the horizon seen directly through the unsilvered half of B, and also the view of it, H, seen by successive reflection at O and the silvered half of B, are coincident. The mirror O is then parallel to B in the position M_1 , and the ray HO is reflected along OB and BD to enter the telescope T. The mirror O is now rotated to a position M_2 until the image of the sun S, seen by successive reflections at O and B, is on the horizon H', and the angle of rotation, θ , of the mirror is noted, Fig. 16.6.

The ray SO from the sun is now reflected in turn from O and B so that it travels along BD, the direction of the horizon, and the angle of deviation of the ray is thus angle SOH. But the angle between the mirrors M_2 and B is θ . Thus, from our result for successive reflections at two inclined mirrors, angle SOH = 2θ . Now the angle of elevation of the sun, S, is angle SOH. Hence the angle of elevation is twice the angle

of rotation of the mirror O , and can thus be easily measured from a scale (not shown) which measures the rotation of O .

Since the angle of deviation after two successive reflections is independent of the angle of incidence on the first mirror (p. 394), the image of the sun S through T will continue to be seen on the horizon once O is adjusted, no matter how the ship pitches or rolls. This is an advantage of the sextant.

Images in Plane Mirrors

So far we have discussed the deviation of light by a plane mirror. We have now to consider the *images* in plane mirrors.

Suppose that a **point object** A is placed in front of a mirror M , Fig. 16.7. A ray AO from A , incident on M , is reflected along OB in such a way that angle $AON = \text{angle } BON$, where ON is the normal at O to the mirror. A ray AD incident normally on the mirror at D is reflected back along DA . Thus the rays reflected from M appear to

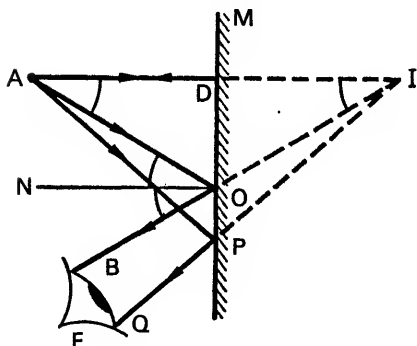


FIG. 16.7. Image in plane mirror.

come from a point I *behind* the mirror, where I is the point of intersection of BO and AD produced. As we shall prove shortly, any ray from A , such as AP , is also reflected as if it comes from I , and hence an observer at E sees the image of A at I .

Since angle $AON = \text{alternate angle } DAO$, and angle $BON = \text{corresponding angle } DIO$, it follows that angle $DAO = \text{angle } DIO$. Thus in the triangles ODA, ODI , angle $DAO = \text{angle } DIO$, OD is common, and angle $ADO = 90^\circ = \text{angle } IDO$. The two triangles are hence congruent, and therefore $AD = ID$. For a given position of the object, A and D are fixed points. Consequently, since $AD = ID$, the point I is a fixed point; and hence *any* ray reflected from the mirror must pass through I , as was stated above.

We have just shown that *the object and image in a plane mirror are at equal perpendicular distances from the mirror*. It should also be noted that $AO = OI$ in Fig. 16.7, and hence the object and image are at equal distances from any point on the mirror.

Image of Finite-sized Object. Perversion

If a right-handed batsman observes his stance in a plane mirror, he appears to be left-handed. Again, the words on a piece of blotting-paper become legible when the paper is viewed in a mirror. This

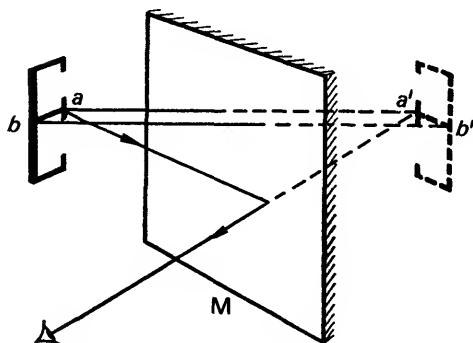


FIG. 16.8. Pervorted (laterally inverted) image.

phenomenon can be explained by considering an E-shaped object placed in front of a mirror M , Fig. 16.8. The image of a point a on the object is at a' at an equal distance behind the mirror, and the image of a point b on the left of a is at b' , which is on the *right* of a' . The left-hand side of the image thus corresponds to the right-hand side of the object, and vice-versa, and the object is said to be *pervorted*, or *laterally inverted* to an observer.

Virtual and Real Images

As was shown on p. 396, an object O in front of a mirror has an image I behind the mirror. The rays reflected from the mirror do not actually pass through I , but only *appear* to do so, and the image cannot be received on a screen because the image is behind the mirror, Fig. 16.9 (i). This type of image is therefore called an unreal or **virtual** image.

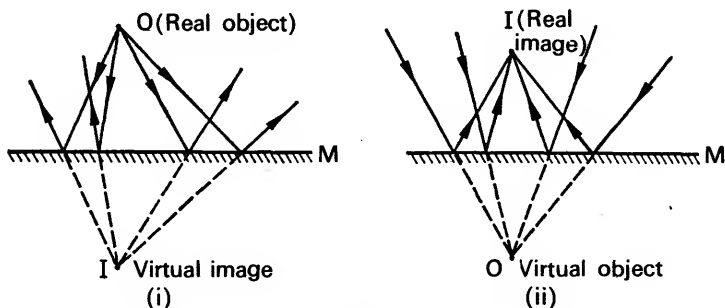


FIG. 16.9. Virtual and real image in plane mirror.

It must not be thought, however, that only virtual images are obtained with a plane mirror. If a *convergent* beam is incident on a plane mirror M , the reflected rays pass through a point I *in front of* M , Fig. 16.9 (ii). If the incident beam converges to the point O , the latter is termed a “virtual” object; I is called a **real** image because it can be received on a

screen. Fig. 16.9 (i) and (ii) should now be compared. In the former, a real object (divergent beam) gives rise to a virtual image; in the latter, a virtual object (convergent beam) gives rise to a real image. In each case the image and object are at equal distances from the mirror.

Location of Images by No Parallax Method

A virtual image can be located by the *method of no parallax*, which we shall now describe.

Suppose O is a pin placed in front of a plane mirror M, giving rise to a virtual image I behind M, Fig. 16.10. A pin P behind the mirror is then

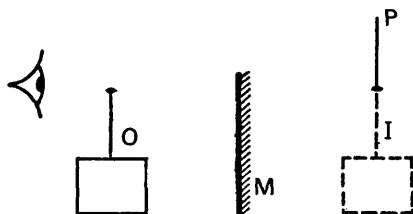


FIG. 16.10. Location of image by no parallax method.

moved towards or away from M, each time moving the head from side to side so as to detect any relative motion between I and P. When the latter appear to move together they are coincident in position, and hence P is at the place of the image I, which is thus located. When P and I do not coincide, they appear to move relative to one another when the observer's head is moved; this relative movement is called "parallax". It is useful to note that the nearer object moves in the opposite direction to the observer.

The method of no parallax can be used, as we shall see later, to locate the positions of real images, as well as virtual images, obtained with lenses and curved mirrors.

Images in Inclined Mirrors

A *kaleidoscope*, produced as a toy under the name of "mirrorscope", consists of two inclined pieces of plane glass with some coloured tinsel between them. On looking into the kaleidoscope a beautiful series of coloured images of the tinsel can often be seen, and the instrument is used by designers to obtain ideas on colouring fashions.

Suppose OA, OB are two mirrors inclined at an angle θ , and P is a point object between them, Fig. 16.11 (i). The image of P in the mirror OB is B_1 , and $OP = OB_1$ (see p. 396). B_1 then forms an image B_2 in the mirror OA, with $OB_2 = OB_1$, B_2 forms an image B_3 in OB, and so on. All the images thus lie on a circle of centre O and radius OP. Another set

of images, $A_1, A_2, A_3 \dots$, have their origin in the image A_1 formed by P in the mirror OA . When the observer looks into the mirror OB he sees

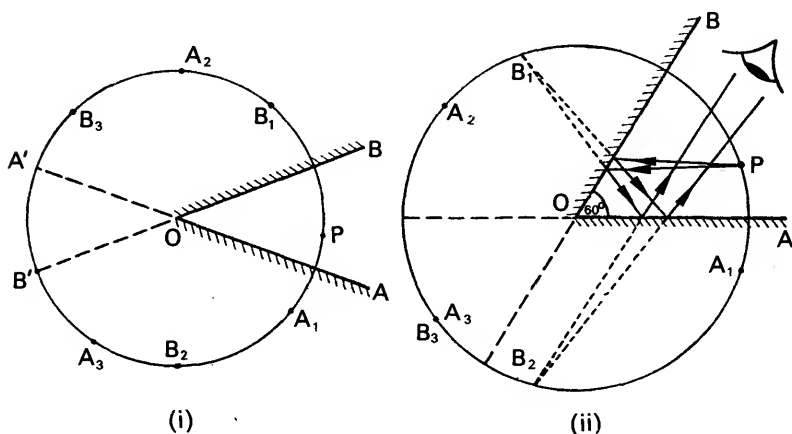


FIG. 16.11. Images in inclined mirrors.

the series of images $B_1, A_2, B_3 \dots$; when he looks into the mirror OA he sees the series of images $A_1, B_2, A_3 \dots$. A finite series of images is seen in either mirror, the last image (not shown) being the one formed on the arc $A'B'$, because it is then *behind* the silvering of the next mirror.

When the mirrors are inclined at an angle of 60° , the final images of P, A_3, B_3 , of each series coincide, Fig. 16.11 (ii). The total number of images is now 5, as the reader can verify. Fig. 16.11 (ii) illustrates the cone of light received by the pupil of the eye when the image B_2 is observed, reflection occurring successively at the mirrors. The drawing is started by joining B_2 to the boundary of the eye, then using B_1 , and finally using P .

EXAMPLE

State the laws of reflection of light and describe how you would verify these laws.

A man 2 m tall, whose eye level is 1.84 m above the ground, looks at his image in a vertical mirror. What is the minimum vertical length of the mirror if the man is to be able to see the whole of himself? Indicate its position accurately in a diagram.

First part. See text. A ray-box can be used to verify the laws.

Second part. Suppose the man is represented by HF , where H is his head and F is his feet; suppose that E represents his eyes, Fig. 16.12. Since the man

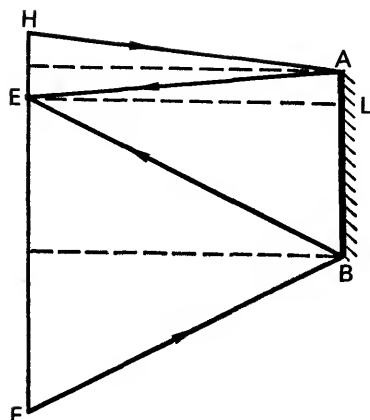


FIG. 16.12. Example.

sees his head H, a ray HA from H to the top A of the mirror is reflected to E. Thus A lies on the perpendicular bisector of HE, and hence $AL = \frac{1}{2} HE = 0.08$ m, where L is the point on the mirror at the same level as E. Since the man sees his feet F, a ray FB from F to the bottom B of the mirror is also reflected to E. Thus the perpendicular bisector of EF passes through B, and hence $BL = \frac{1}{2} FE = \frac{1}{2} \times 1.84$ m = 0.92 m.

\therefore length of mirror = $AL + LB = 0.08$ m + 0.92 m = 1 m.

EXERCISES 16

1. Prove the relation between the angle of rotation of a mirror and the angle of deflection of a reflected ray, when the direction of the incident ray is constant.

2. Two plane mirrors are inclined at an angle of 35° . A ray of light is incident on one mirror at 60° , and undergoes two successive reflections at the mirrors. Show by accurate drawing that the angle of deviation produced is 70° .

Repeat with an angle of incidence of 45° , instead of 60° , and state the law concerning the angle of deviation.

3. Two plane mirrors are inclined to each other at a fixed angle. If a ray travelling in a plane perpendicular to both mirrors is reflected first from one and then from the other, show that the angle through which it is deflected does not depend on the angle at which it strikes the first mirror.

Describe and explain the action of *either* a sextant *or* a rear reflector on a bicycle. (L.)

4. State the laws of reflection of light. Two plane mirrors are parallel and face each other. They are a cm apart and a small luminous object is placed b cm from one of them. Find the distance from the object of an image produced by four reflections. Deduce the corresponding distance for an image produced by $2n$ reflections. (L.)

5. Two vertical plane mirrors A and B are inclined to one another at an angle a . A ray of light, travelling horizontally, is reflected first from A and then from B. Find the resultant deviation and show it is independent of the original direction of the ray. Describe an optical instrument that depends on the above proposition. (N.)

6. State the laws of reflection for a parallel beam of light incident upon a plane mirror.

Indicate clearly by means of diagrams (a) how the position and size of the image of an extended object may be determined by geometrical construction, in the case of reflection in a plane; (b) how the positions of the images of a small lamp, placed unsymmetrically between parallel reflecting planes, may be graphically determined. (W.)

7. Describe the construction of the *sextant* and the *periscope*. Illustrate your answer by clear diagrams and indicate the optical principles involved. (L.)

chapter seventeen

Reflection at curved mirrors

CURVED mirrors are reputed to have been used thousands of years ago. Today motor-cars and other vehicles are equipped with driving mirrors which are curved, searchlights have curved mirrors inside them, and the largest telescope in the world utilises a huge curved mirror (p. 544).

Convex and Concave Mirrors. Definitions

In the theory of Light we are mainly concerned with curved mirrors which are parts of *spherical* surfaces. In Fig. 17.1 (a), the mirror APB is part of a sphere whose centre C is in front of the reflecting surface; in Fig. 17.1 (b), the mirror KPL is part of a sphere whose centre C is behind its reflecting surface. To a person in front of it APB curves inwards and is known as a **concave mirror**, while KPL bulges outwards and is known as a **convex mirror**.

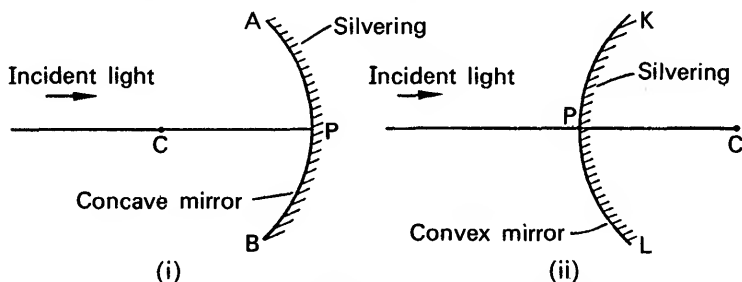


FIG. 17.1. Concave (converging) and convex (diverging) mirrors.

The mid-point, P, of the mirror is called its *pole*; C, the centre of the sphere of which the mirror is part, is known as the *centre of curvature*; and AB is called the *aperture* of the mirror. The line PC is known as the *principal axis*, and plays an important part in the drawing of images in the mirrors; lines parallel to PC are called *secondary axes*.

Narrow and Wide Beams. The Caustic

When a very narrow beam of rays, parallel to the principal axis and close to it, is incident on a concave mirror, experiment shows that all the reflected rays converge to a point F on the principal axis, which is therefore known as the *principal focus* of the mirror, Fig. 17.2 (i). On this account a concave mirror is better described as a “converging”

mirror. An image of the sun, whose rays on the earth are parallel, can hence be received on a screen at F , and thus a concave mirror has a *real* focus.

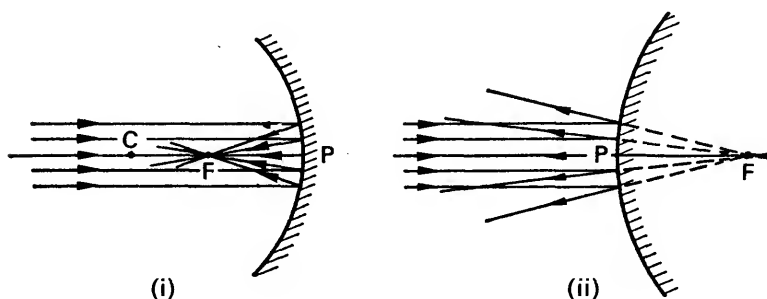


FIG. 17.2. Foci of concave and convex mirrors.

If a narrow beam of parallel rays is incident on a convex mirror, experiment shows that the reflected rays form a divergent beam which appear to come from a point F *behind* the mirror, Fig. 17.2 (ii). A convex mirror has thus a *virtual* focus, and the image of the sun cannot be received on a screen using this type of mirror. To express its action on a parallel beam of light, a convex mirror is often called a “diverging” mirror.

When a *wide* beam of light, parallel to the principal axis, is incident on a concave spherical mirror, experiment shows that reflected rays do not pass through a single point, as was the case with a narrow beam. The reflected rays appear to touch a surface known as a *caustic surface*, S , which has an apex, or cusp, at F , the principal focus, Fig. 17.3. Similarly, if a wide beam of parallel light is incident on a convex mirror, the reflected rays do not appear to diverge from a single point, as was the case with a narrow beam.

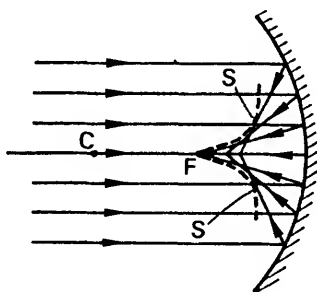


FIG. 17.3. Caustic surface.

Parabolic Mirrors

If a small lamp is placed at the focus, F , of a concave mirror, it follows from the principle of the reversibility of light (p. 390) that rays striking the mirror round a small area about the pole are reflected parallel. See Fig. 17.2 (i). But those rays from the lamp which strike the mirror at points well away from P will be reflected in different directions, because a *wide* parallel beam is not brought to a focus at F , as shown in Fig. 17.3. The beam of light reflected from the mirror thus diminishes

This relation between f and r is the same for the case of the convex mirror, Fig. 17.5 (ii), as the reader can easily verify.

Images in Concave Mirrors

Concave mirrors produce images of different sizes; sometimes they are inverted and real, and on other occasions they are erect (the same way up as the object) and virtual. As we shall see, the nature of the image formed depends on the distance of the object from the mirror.

Consider an object of finite size OH placed at O perpendicular to the principal axis of the mirror, Fig. 17.6 (i). The image, R , of the top point

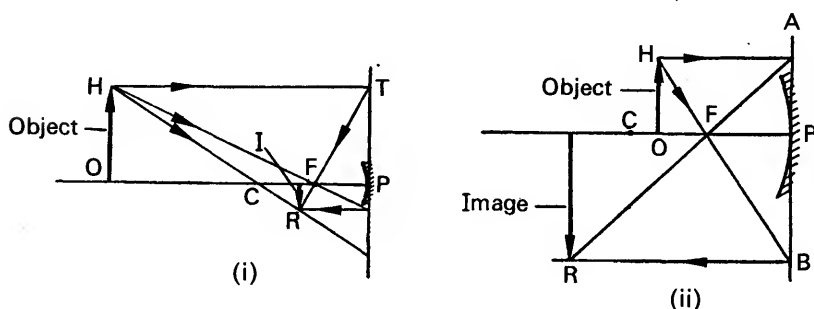


FIG. 17.6. Images in concave mirrors.

H can be located by the intersection of two reflected rays coming initially from H , and the rays usually chosen are two of the following: (1) The ray HT parallel to the principal axis, which is reflected to pass through the focus, F , (2) the ray HC passing through the centre of curvature, C , which is reflected back along its own path because it is a normal to the mirror, (3) the ray HF passing through the focus, F , which is reflected parallel to the principal axis. Since the mirror has a small aperture, and we are considering a narrow beam of light, the mirror must be represented in accurate image drawings by a *straight* line. Thus PT in Fig. 17.6 (i) represents a perfect mirror.

When the object is a very long distance away (at infinity), the image is small and is formed inverted at the focus (p. 402). As the object approaches the centre of curvature, C , the image remains real and inverted, and is formed in front of the object, Fig. 17.6 (i). When the object is between C and F , the image is real, inverted, and larger than the object; it is now further from the mirror than the object, Fig. 17.6 (ii).

As the object approaches the focus, the image recedes further from the mirror, and when the object is at the focus, the image is at infinity. When the object is nearer to the mirror than the focus the image IR becomes *erect* and *virtual*, as shown in Fig. 17.7 (i). In this case the image

is *magnified*, and the concave mirror can thus be used as a shaving mirror.

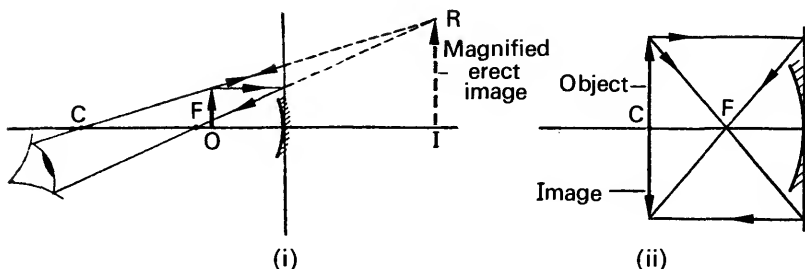


FIG. 17.7. Images in concave mirrors.

A *special case* occurs when the object is at the centre of curvature, C . The image is then real, inverted, and the same size of the object, and it is also situated at C , Fig. 17.7 (ii). This case provides a simple method of locating the centre of curvature of a concave mirror (p. 413).

Images in Convex Mirrors

Experiment shows that the image of an object in a convex mirror is erect, virtual, and diminished in size, no matter where the object is situated. Suppose an object OH is placed in front of a convex mirror, Fig. 17.8 (i). A ray HM parallel to the principal axis is reflected as if it

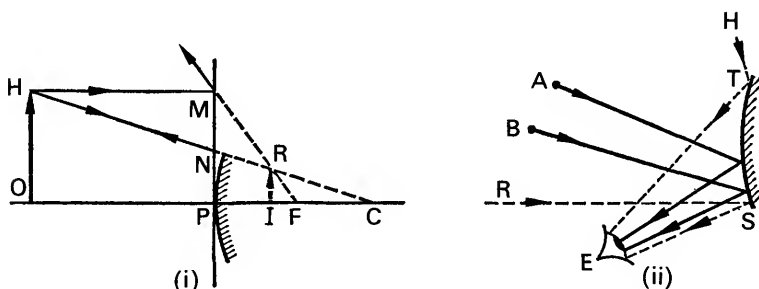


FIG. 17.8. Images in convex mirrors.

appeared to come from the virtual focus, F , and a ray HN incident towards the centre of curvature, C , is reflected back along its path. The two reflected rays intersect *behind* the mirror at R , and IR is a virtual and erect image.

Objects well outside the principal axis of a convex mirror, such as A , B in Fig. 17.8 (ii), can be seen by an observer at E , whose *field of view* is that between HT and RS , where T , S are the edges of the mirror. Thus in addition to providing an erect image the convex mirror has a wide field of view, and is hence used as a driving mirror.

Formulae for Mirrors. Sign Convention

Many of the advances in the uses of curved mirrors and lenses have resulted from the use of *optical formulae*, and we have now to consider the relation which holds between the object and image distances in mirrors and their focal length. In order to obtain a formula which holds for both concave and convex mirrors, a *sign rule* or *convention* must be obeyed, and we shall adopt the following:

A real object or image distance is a positive distance.

A virtual object or image distance is a negative distance.

In brief, "real is positive, virtual is negative". The focal length of a concave mirror is thus a positive distance; the focal length of a convex mirror is a negative distance.

Concave Mirror

Consider a point object O on the principal axis of a concave mirror. A ray OX from O is reflected in the direction XI making an equal angle θ with the normal CX; a ray OP from O, incident at P, is reflected back along PO, since CP is the normal at P. The point of intersection, I, of the two rays is the image of O. Fig. 17.9.

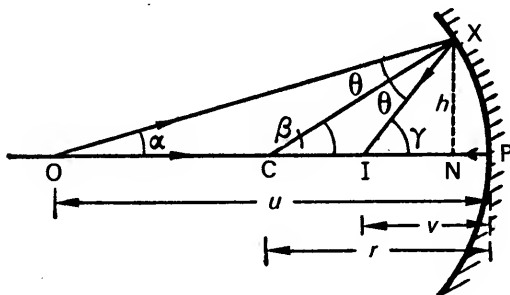


FIG. 17.9. Mirror formula.

Suppose α , β , γ are the angles made by OX, CX, IX respectively with the axis. Since we are considering a mirror of small aperture these angles are small in practice, Fig. 17.9 being exaggerated. As β is the exterior angle of triangle CXO, we have $\beta = \alpha + \theta$.

$$\therefore \theta = \beta - \alpha \quad \dots \quad (i)$$

Since γ is the exterior angle of triangle IXC, we have $\gamma = \beta + \theta$.

$$\therefore \theta = \gamma - \beta \quad \dots \quad (ii)$$

From (i) and (ii), it follows that

$$\begin{aligned} \beta - \alpha &= \gamma - \beta \\ \therefore \alpha + \gamma &= 2\beta \quad \dots \quad (iii) \end{aligned}$$

We can now substitute for α , β , γ in terms of h , the height of X above the axis, and the distances OP, CP, IP. In so doing (a) we assume N is practically coincident with P, as X is very close to P in practice, (b) the appropriate sign, + or - , must precede all the numerical values of the distances concerned. Also, as $\alpha = \tan \alpha$ in radians when α is very small, we have

$$\alpha = \frac{XN}{ON} = \frac{XN}{+OP} = \frac{h}{+OP},$$

where OP is the distance of the real object O from the mirror in centimetres, say, and $XN = h$. Similarly,

$$\beta = \frac{XN}{CN} = \frac{XN}{+CP} = \frac{h}{+CP},$$

as CP, the radius of curvature of the concave mirror, is real.

Also,
$$\gamma = \frac{XN}{IN} = \frac{XN}{+IP} = \frac{h}{+IP},$$

where IP is the distance of the real image I from the mirror. Substituting for α , β , γ in (iii),

$$\frac{h}{(+IP)} + \frac{h}{(+OP)} = 2 \frac{h}{(+CP)}.$$

Dividing by h ,
$$\frac{1}{IP} + \frac{1}{OP} = \frac{2}{CP}.$$

If we let v represent the image distance from the mirror, u the object distance from the mirror, and r the radius of curvature, we have

$$\frac{1}{v} + \frac{1}{u} = \frac{2}{r} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Further, since $f = r/2$, then $\frac{2}{r} = \frac{1}{f}$.

$$\therefore \frac{1}{v} + \frac{1}{u} = \frac{1}{f} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The relations (2), (3) are general formulae for curved spherical mirrors; and when they are used the appropriate sign for v , u , f , or r must always precede the corresponding numerical value.

Convex Mirror

We now obtain a relation for object distance (u), image distance (v), and focal length (f) of a convex mirror. In this case the incident rays OX, OP are reflected as if they appear to come from the point I behind the mirror, which is therefore a virtual image, and hence the image

distance IP is *negative*, Fig. 17.10. Further, CP is *negative*, as the centre of curvature of X , a convex mirror, is behind the mirror.

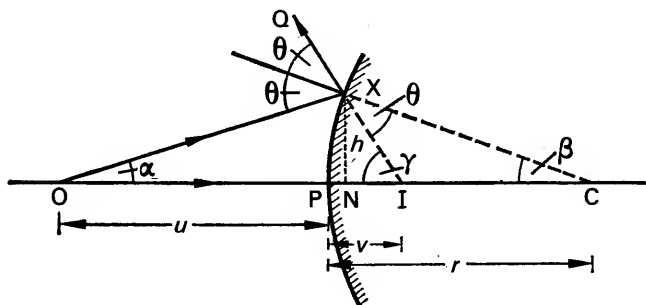


FIG. 17.10. Mirror formula.

Since θ is the exterior angle of triangle COX , $\theta = \alpha + \beta$. As γ is the exterior angle of triangle CIX , $\gamma = \theta + \beta$, or $\theta = \gamma - \beta$.

$$\therefore \gamma - \beta = \alpha + \beta$$

$$\therefore \gamma - \alpha = 2\beta \quad \dots \quad (i)$$

Now $\gamma = \frac{h}{IN} = \frac{h}{(-IP)}$, as I is virtual; $\alpha = \frac{h}{ON} = \frac{h}{(+OP)}$, as O is real,

$\beta = \frac{h}{NC} = \frac{h}{(-PC)}$, as C is virtual. Substituting in (i),

$$\therefore \frac{h}{(-IP)} - \frac{h}{(+OP)} = \frac{2h}{(-CP)}$$

$$\therefore \frac{1}{IP} + \frac{1}{OP} = \frac{2}{CP}$$

$$\therefore \frac{1}{v} + \frac{1}{u} = \frac{2}{r}$$

and

$$\therefore \frac{1}{v} + \frac{1}{u} = \frac{1}{f}.$$

Thus, using the sign convention, the same formula holds for concave and convex mirrors.

Formula for Magnification

The lateral magnification, m , produced by a mirror is defined by

$$m = \frac{\text{height of image}}{\text{height of object}}$$

Suppose IR is the image of an object OH in a concave or convex mirror. Fig. 17.11 (i), (ii). Then a ray HP from the top point H of the

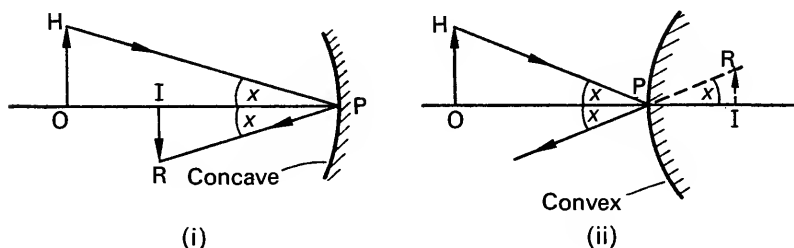


FIG. 17.11. Magnification formula.

object passes through the top point R of the image after reflection from the mirror. Now the normal to the mirror at P is the principal axis, OP. Thus angle OPH = angle IPR, from the law of reflection.

$$\therefore \tan OPH = \tan IPR$$

$$\text{i.e., } \frac{OH}{OP} = \frac{IR}{IP}$$

$$\therefore \frac{IR}{OH} = \frac{IP}{OP}$$

But IP = image distance = v , and OP = object distance = u

$$\therefore \frac{IR}{OH} = \frac{v}{u}$$

$$\therefore m = \frac{v}{u} \quad (4)$$

Thus if the image distance is half the object distance, the image is half the length of the object.

Since

$$\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$$

$$\therefore \frac{v}{v} + \frac{v}{u} = \frac{v}{f},$$

multiplying throughout by v .

$$1 + \frac{v}{u} = \frac{v}{f}$$

$$\therefore 1 + m = \frac{v}{f}$$

$$\therefore m = \frac{v}{f} - 1$$

Some Applications of Mirror Formulae

The following examples will assist the reader to understand how to apply the formulae $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$ and $m = \frac{v}{u}$ correctly:

1. An object is placed 10 cm in front of a concave mirror of focal length 15 cm. Find the image position and the magnification.

Since the mirror is concave, $f = +15$ cm. The object is real, and hence

$u = +10$ cm. Substituting in $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$,

$$\begin{aligned}\frac{1}{v} + \frac{1}{(+10)} &= \frac{1}{(+15)} \\ \therefore \frac{1}{v} &= \frac{1}{15} - \frac{1}{10} = -\frac{1}{30} \\ \therefore v &= -30\end{aligned}$$

Since v is negative in sign the image is *virtual*, and it is 30 cm from the mirror. See Fig. 17.7 (i). The magnification, $m = \frac{v}{u} = \frac{30}{10} = 3$, so that the image is three times as high as the object.

2. The image of an object in a convex mirror is 4 cm from the mirror. If the mirror has a radius of curvature of 24 cm, find the object position and the magnification.

The image in a convex mirror is always virtual (p. 406). Hence $v = -4$ cm. The focal length of the mirror $= \frac{1}{2}r = 12$ cm; and since the mirror is convex, $f = -12$ cm. Substituting in $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$

$$\begin{aligned}\frac{1}{(-4)} + \frac{1}{u} &= \frac{1}{(-12)} \\ \therefore \frac{1}{u} &= -\frac{1}{12} + \frac{1}{4} = \frac{1}{6} \\ \therefore u &= 6\end{aligned}$$

Since u is positive in sign the object is real, and it is 6 cm from the mirror.

The magnification, $m = \frac{v}{u} = \frac{4}{6} = \frac{2}{3}$, and hence the image is two-thirds as high as the object. See Fig. 17.8 (i).

3. An erect image, three times the size of the object, is obtained with a concave mirror of radius of curvature 36 cm. What is the position of the object?

If x cm is the numerical value of the distance of the object from the mirror,

the image distance must be $3x$ cm, since the magnification $m = \frac{\text{image distance}}{\text{object distance}} = 3$. Now an *erect* image is obtained with a concave mirror only when the image is *virtual* (p. 406).

$$\therefore \text{image distance, } v = -3x$$

$$\text{Also, object distance, } u = +x$$

$$\text{and focal length, } f = \frac{1}{2}r = +18 \text{ cm.}$$

Substituting in $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$,

$$\therefore \frac{1}{(-3x)} + \frac{1}{(+x)} = \frac{1}{(+18)}$$

$$\therefore -\frac{1}{3x} + \frac{1}{x} = \frac{1}{18}$$

$$\therefore \frac{2}{3x} = \frac{1}{18}$$

$$\therefore x = 12$$

Thus the object is 12 cm from the mirror.

Virtual Object and Convex Mirror

We have already seen that a convex mirror produces a virtual image of an object in front of it, which is a real object. A convex mirror may sometimes produce a real image of a *virtual* object.

As an illustration, consider an incident beam of light bounded by AB, DE, converging to a point O *behind* the mirror, Fig. 17.12. O is

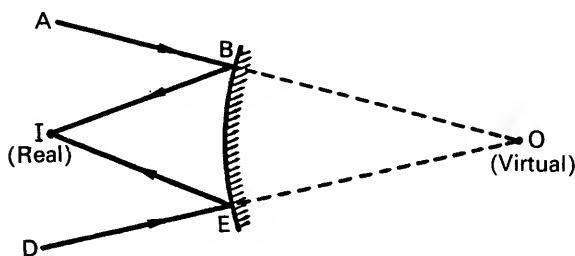


FIG. 17.12. Real image in convex mirror.

regarded as a virtual object, and if its distance from the mirror is 10 cm, then the object distance, $u = -10$. Suppose the convex mirror has a focal length of 15 cm, i.e., $f = -15$.

Since

$$\frac{1}{v} + \frac{1}{u} = \frac{1}{f},$$

$$\frac{1}{v} + \frac{1}{(-10)} = \frac{1}{(-15)}$$

$$\therefore \frac{1}{v} = -\frac{1}{15} + \frac{1}{10} = +\frac{1}{30}$$

$$\therefore v = +30$$

The point image, I, is thus 30 cm from the mirror, and is *real*. The beam reflected from the mirror is hence a convergent beam, Fig. 17.12; a similar case with a plane mirror is shown in Fig. 16.9 (ii).

Object at Centre of Curvature of Concave Mirror

Suppose an object is placed at the centre of curvature of a concave mirror. Then $u = +r$, where r is the numerical value of the radius of

curvature. Substituting in $\frac{1}{v} + \frac{1}{u} = \frac{2}{r}$ to find the image distance v ,

$$\frac{1}{v} + \frac{1}{(+r)} = \frac{2}{(+r)}$$

$$\therefore \frac{1}{v} = \frac{2}{r} - \frac{1}{r} = \frac{1}{r}$$

$$\therefore v = r$$

The image is therefore also formed at the centre of curvature. The

magnification in this case is given by $m = \frac{v}{u} = \frac{r}{r} = 1$, and hence the

object and image are the same size. This case is illustrated in Fig. 17.7 (ii), to which the reader should now refer.

SOME METHODS OF DETERMINING FOCAL LENGTH AND RADIUS OF CURVATURE OF MIRRORS

Concave Mirror

Method 1. A pin O is placed above a concave mirror M so that an inverted image of the pin can be seen, Fig. 17.13. If the pin is moved up and down with its point on the axis of the mirror, and an observer E moves his eye perpendicularly to the pin at the same time, a position of O is reached when the image I remains perfectly in line with O as E moves; i.e., there is no parallax (no relative displacement) between

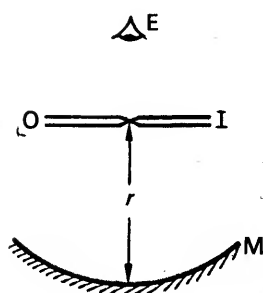


FIG. 17.13. Centre of curvature of concave mirror.

pin and image. The pin is now at exactly the same place as its image I, Fig. 17.13. Since an object and image coincide in position at the centre of curvature of a concave mirror (p. 413), the distance from the point of the pin to the mirror is equal to its radius of curvature, r . The focal length, f , which is $\frac{r}{2}$, is then easily obtained.

If an illuminated object is available instead of a pin, the object is moved to or from the mirror until a clear image is obtained beside the object. The distance of the object from the mirror is then equal to r .

In general the method of no parallax, using a pin, gives a higher degree of accuracy in locating an image.

Method 2. By using the method of no parallax, or employing an illuminated object, several, say six, values of the image distance, v , can be obtained with the concave mirror, corresponding to six different values of the object distance u . Substituting for u , v in the formula $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, six values of f can be calculated, and the average value taken.

A better method of procedure, however, is to plot the magnitudes of $\frac{1}{v}$ against $\frac{1}{u}$; a straight line BA can be drawn through the points thus obtained, Fig. 17.14. Now $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$. Hence, when $\frac{1}{v} = 0$, $\frac{1}{u} = \frac{1}{f}$.

But
$$OB = \frac{1}{u} \text{ when } \frac{1}{v} = 0$$

$$\therefore OB = \frac{1}{f}, \text{ i.e., } f = \frac{1}{OB}.$$

Thus the focal length can be determined from the reciprocal value of the intercept OB on the axis of $\frac{1}{u}$.

From $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, $\frac{1}{v} = \frac{1}{f}$ when $\frac{1}{u} = 0$. Thus $OA = \frac{1}{f}$, Fig. 17.14, and hence $f = \frac{1}{OA}$. It can thus be seen that (i) f can also be calculated from

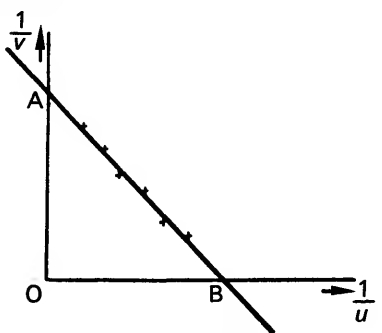


FIG. 17.14. Graph of $\frac{1}{v}$ against $\frac{1}{u}$.

the reciprocal of the intercept OA on the axis of $\frac{1}{v}$, (ii) OAB is an isosceles triangle if the same scale is employed for $\frac{1}{v}$ and $\frac{1}{u}$.

Convex Mirror

Method 1. By using a convex lens, L , a real image of an object O can be formed at a point C on the other side of L , Fig. 17.15. The convex mirror, MN , is then placed between L and C with its reflecting face facing the lens, so that a convergent beam of rays is incident on the mirror. When the latter is moved along the axis OC a position will be reached when the beam is incident *normally* on the mirror, in which case the rays are reflected back along the incident path. A *real inverted image is then formed at O* .

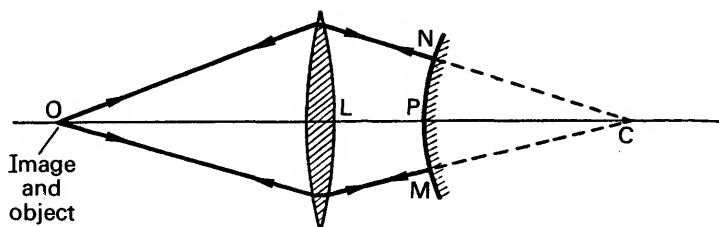


FIG. 17.15. Convex mirror measurement.

Since the rays incident on the mirror, for example at N or M , are normal to the mirror surface, they will, if produced, pass through the centre of curvature, C , of the mirror. Thus the distance $PC = r$, the radius of curvature. Since $PC = LC - LP$, this distance can be obtained from measurement of LP and LC , the latter being the image distance from the lens when the mirror is taken away.

Method 2. A more difficult method than the above consists of positioning a pin O in front of the convex mirror, when a virtual image, I , is formed, Fig. 17.16. A small plane mirror M is then moved between O and P until the image I' of the lower part of O in M coincides in position with the upper part of the image of O in the convex mirror. The distances OP , MP are then measured.

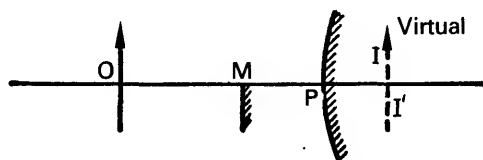


FIG. 17.16. Convex mirror measurement.

Since M is a plane mirror, the image I' of O in it is such that $OM = MI'$. Thus $PI = MI' - MP = OM - MP$, and hence PI can be calculated. But $PI = v$, the image distance of O in the convex mirror, and $OP = u$, the object distance. Substituting for the virtual distance v , and u , in $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, the focal length of the convex mirror can be found.

EXAMPLES

1. Show that a concave spherical mirror can produce a focused image of an object when certain conditions are observed, and prove the usual relation between the object and image distances. A linear object, 10 cm long, lies along the axis of a concave mirror whose radius of curvature is 30 cm, the near end of the object lying 18 cm from the mirror. Find the magnification of the image. (W.)

First part. The condition for a focused image is that the light from the object must be incident as a narrow beam round the pole of the mirror. This implies that the object must be small (p. 402). The usual relation between the object and image distances is proved on p. 407.

Second part. Suppose Q is the near end of the object, which is 18 cm from the mirror. The distance of the image of Q is given by

$$\frac{1}{v} + \frac{1}{(+18)} = \frac{1}{(+15)}$$

$$\text{since } f = \frac{r}{2} = \frac{30}{2} = 15 \text{ cm}$$

$$\therefore \frac{1}{v} = \frac{1}{15} - \frac{1}{18}$$

from which

$$v = 90 \text{ cm}$$

The other end P of the object is $(10 + 18)$ cm from the mirror, or 28 cm. The image of P is given by

$$\frac{1}{v} + \frac{1}{(+28)} = \frac{1}{(+15)}$$

$$\therefore \frac{1}{v} = \frac{1}{15} - \frac{1}{28}$$

from which

$$v = 32.3 \text{ cm}$$

$$\therefore \text{length of image} = 90 - 32.3 = 57.7 \text{ cm}$$

$$\therefore \text{magnification of image} = \frac{57.7}{10} = 5.8$$

2. PBCA is the axis of a concave spherical mirror, A being a point object, B its image, C the centre of curvature of the mirror and P the pole. Find a relation between PA, PB, and PC, supposing the aperture of the mirror to be small. A concave mirror forms, on a screen, a real image of twice the linear

dimensions of the object. Object and screen are then moved until the image is three times the size of the object. If the shift of the screen is 25 cm determine the shift of the object and the focal length of the mirror. (N.)

First part. See text.

Second part. Suppose v is the distance of the screen from the mirror when the image is twice the length of the object. Since the magnification, m , is given by

$$m = \frac{v}{f} - 1 \quad . \quad . \quad . \quad . \quad (\text{i})$$

where f is the focal length of the mirror (see p. 410),

$$\therefore 2 = \frac{v}{f} - 1 \quad \dots \dots \dots \text{(ii)}$$

When $m = 3$, the image distance is $(v + 25)$ cm. Substituting in (i),

$$\therefore 3 = \frac{v + 25}{f} - 1 \quad \dots \dots \dots \text{(iii)}$$

Subtracting (ii) from (iii), we have

$$1 = \frac{25}{f}$$
$$\therefore f = 25 \text{ cm.}$$

From (ii), $2 = \frac{v}{25} - 1$, or $v = 75$ cm. The object distance, u , is thus given by $\frac{1}{75} + \frac{1}{u} = \frac{1}{25}$, from which $u = 37\frac{1}{2}$ cm. From (iii), $v = 100$ cm. The object distance, u , is then given by $\frac{1}{100} + \frac{1}{u} = \frac{1}{25}$, from which $u = 33\frac{1}{3}$ cm. Thus the shift of the object $= 37\frac{1}{2} - 33\frac{1}{3} = 4\frac{1}{6}$ cm.

EXERCISES 17

1. An object is placed (i) 10 cm, (ii) 4 cm from a concave mirror of radius of curvature 12 cm. Calculate the image position in each case, and the respective magnifications.
2. Repeat Q. 1 by accurate drawings to scale. (*Note.*—The mirror must be represented by a straight line.)
3. An object is placed 15 cm from a convex mirror of focal length 10 cm. Calculate the image distance and the magnification produced. Draw an accurate diagram to scale, and verify your drawing from the calculated results.
4. Explain with the aid of diagrams why a curved mirror can be used (i) as a driving mirror, (ii) in a searchlight, (iii) as a shaving mirror. Why is a special form of mirror required in the searchlight?

5. Describe and explain a method of finding the focal length of (a) a concave mirror, (b) a convex mirror.

6. A pole 4 m long is laid along the principal axis of a convex mirror of focal length 1 m. The end of the pole nearer the mirror is 2 m from it. Find the length of the image of the pole.

7. Deduce a formula connecting u , v and r , the distances of object, image and centre of curvature from a spherical mirror.

A mirror forms an erect image 30 cm from the object and twice its height. Where must the mirror be situated? What is its radius of curvature? Assuming the object to be real, determine whether the mirror is convex or concave. (L.)

8. Establish the formula $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$ for a concave mirror.

In an experiment with a concave mirror the magnification m of the image is measured for a series of values of v , and a curve is plotted between m and v . What curve would you expect to obtain, and how would you use it to deduce the focal length of the mirror? (C.)

9. Derive an approximate relation connecting the distances of an object and its image from the surface of a convex spherical mirror.

A small object is placed at right angles to the axis of a concave mirror so as to form (a) a real, (b) a virtual image, twice as long as the object. If the radius of curvature of the mirror is R what is the distance between the two images? (L.)

10. Deduce a formula connecting the distances of object and image from a spherical mirror. What are the advantages of a concave mirror over a lens for use in an astronomical telescope?

A driving mirror consists of a cylindrical mirror of radius 10 cm and length (over the curved surface) of 10 cm. If the eye of the driver be assumed at a great distance from the mirror, find the angle of view. (O. & C.)

11. Find the relation connecting the focal length of a convex spherical mirror with the distances from the mirror of a small object and the image formed by the mirror.

A convex mirror, radius of curvature 30 cm, forms a real image 20 cm from its surface. Explain how this is possible and find whether the image is erect or inverted. (L.)

12. What conditions must be satisfied for an optical system to form an image of an object? Show how these conditions are satisfied for a convex spherical mirror when a small object is placed on its axis and derive a relationship showing how the position of the image depends on the position of the object and the radius of curvature of the mirror.

A millimetre scale is placed at right angles to the axis of a convex mirror of radius of curvature 12 cm. This scale is 18 cm away from the pole of the mirror. Find the position of the image of the scale. What is the size of the divisions of the image? What is the ratio of the angle subtended by the image to that subtended by the object at a point on the axis 25 cm away from the object on the side remote from the mirror? (O. & C.)

13. Describe an experiment to determine the radius of curvature of a convex mirror by an optical method. Illustrate your answer with a ray diagram and explain how the result is derived from the observations.

A small convex mirror is placed 60 cm from the pole and on the axis of a large concave mirror, radius of curvature 200 cm. The position of the convex mirror is such that a real image of a distant object is formed in the plane of a hole drilled through the concave mirror at its pole. Calculate (a) the radius of curvature of the convex mirror, (b) the height of the real image if the distant object subtends an angle of 0.50° at the pole of the concave mirror. Draw a ray diagram to illustrate the action of the convex mirror in producing the image of a non-axial point of the object and suggest a practical application of this arrangement of mirrors. (N.)

chapter eighteen

Refraction at plane surfaces

Laws of Refraction

WHEN a ray of light AO is incident at O on the plane surface of a glass medium, observation shows that some of the light is reflected from the surface along OC in accordance with the laws of reflection, while the rest of the light travels along a new direction, OB, in the glass, Fig. 18.1.

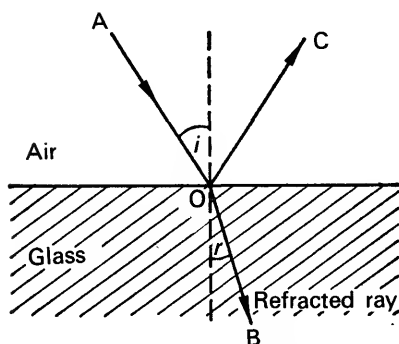


FIG. 18.1. Refraction at plane surface.

On account of the change in direction the light is said to be “refracted” on entering the glass; and the *angle of refraction*, r , is the angle made by the refracted ray OB with the normal at O.

Historical records reveal that the astronomer PTOLEMY, who lived about A.D. 140, measured numerous values of the angle of incidence, i , and the angle of refraction, r , for glass as the angle of incidence was varied. However, he was unable to discover any relation between i and r .

Later scientists were equally unsuccessful, until centuries later SNELL, a Dutch professor, discovered in 1620 that the sines of the angles bear a constant ratio to each other. The *laws of refraction* are:

1. *The incident and refracted rays, and the normal at the point of incidence, all lie in the same plane.*

2. *For two given media, $\frac{\sin i}{\sin r}$ is a constant, where i is the angle of incidence and r is the angle of refraction (Snell's law).*

Refractive Index.

The constant ratio $\frac{\sin i}{\sin r}$ is known as the **refractive index** for the two given media; and as the magnitude of the constant depends on the colour of the light, it is usually specified as that obtained for yellow light. If the medium containing the incident ray is denoted by 1, and that containing the refracted ray by 2, the refractive index can be denoted by n_2 .

Scientists have drawn up tables of refractive indices when the incident

ray is travelling *in vacuo* and is then refracted into the medium concerned, for example, glass or water. The values thus obtained are known as the *absolute* refractive indices of the media; and as a vacuum is always the first medium, the subscripts for the absolute refractive index, symbol n , can be dropped. The magnitude of n for glass is about 1.5, n for water is about 1.33, and n for air at normal pressure is about 1.00028. As the magnitude of the refractive index of a medium is only very slightly altered when the incident light is in air instead of a vacuum, experiments to determine the absolute refractive index n are usually performed with the light incident from air on to the medium; thus we can take $\text{air}n_{\text{glass}}$ as equal to $\text{vacuum}n_{\text{glass}}$ for most practical purposes.

We have already mentioned that light is refracted because it has different velocities in different media. The Wave Theory of Light, discussed on p. 679, shows that the refractive index ${}_1n_2$ for two given media 1 and 2 is given by

$${}_1n_2 = \frac{\text{velocity of light in medium 1}}{\text{velocity of light in medium 2}} \quad (1)$$

and this is a *definition* of refractive index which can be used instead of the ratio $\frac{\sin i}{\sin r}$. An alternative definition of the absolute refractive index, n , of a medium is thus

$$n = \frac{\text{velocity of light in a vacuum}}{\text{velocity of light in medium}} \quad (2)$$

In practice the velocity of light in air can replace the velocity in *vacuo* in this definition.

Relations Between Refractive Indices

(1) Consider a ray of light, AO, refracted from *glass* to *air* along the direction OB; observation then shows that the refracted ray OB is bent away from the normal, Fig. 18.2. The refractive index from glass to air, ${}_g n_a$, is given by $\sin x / \sin y$, by definition, where x is the angle of incidence in the glass and y is the angle of refraction in the air.

From the principle of the reversibility of light (p. 390), it follows that a ray travelling along BO in air is refracted along OA in the glass. The refractive index from air to glass, ${}_a n_g$, is then given by $\sin y / \sin x$, by definition. But ${}_g n_a = \frac{\sin x}{\sin y}$, from the previous paragraph.

$$\therefore {}_g n_a = \frac{1}{{}_a n_g} \quad (3)$$

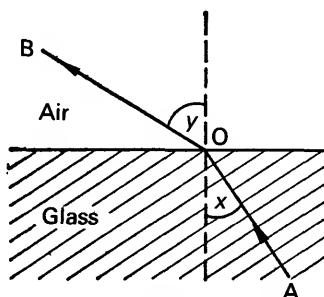


FIG. 18.2. Refraction from glass to air.

If ${}_a n_g$ is 1.5, then ${}_g n_a = \frac{1}{1.5} = 0.67$. Similarly, if the refractive index from air to water is $4/3$, the refractive index from water to air is $3/4$.

(2) Consider a ray AO incident in air on a plane glass boundary, then refracted from the glass into a water medium, and finally emerging along a direction CD into air. *If the boundaries of the media are parallel, experiment shows that the emergent ray CD is parallel to the incident ray AO*, although there is a relative displacement. Fig. 18.3. Thus the angles made with the normals by AO, CD are equal, and we shall denote them by i_a .

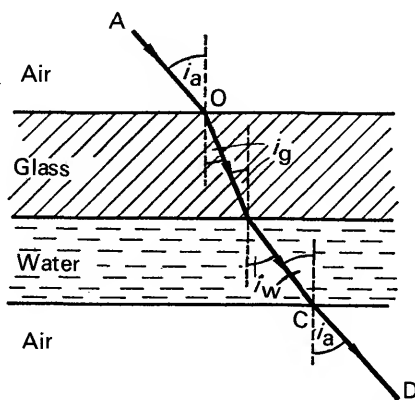


FIG. 18.3. Refraction at parallel plane surfaces.

Suppose i_g, i_w are the angles made with the normals by the respective rays in the glass and water media. Then, by definition, ${}_g n_w = \frac{\sin i_g}{\sin i_w}$.

But
$$\frac{\sin i_g}{\sin i_w} = \frac{\sin i_g}{\sin i_a} \times \frac{\sin i_a}{\sin i_w},$$

and
$$\frac{\sin i_g}{\sin i_a} = {}_g n_a, \text{ and } \frac{\sin i_a}{\sin i_w} = {}_a n_w$$

$$\therefore {}_g n_w = {}_g n_a \times {}_a n_w \quad \dots \dots \dots (i)$$

Further, as ${}_g n_a = \frac{1}{{}_a n_g}$, we can write

$${}_g n_w = \frac{{}_a n_w}{{}_a n_g}.$$

Since ${}_a n_w = 1.33$ and ${}_a n_g = 1.5$, it follows that ${}_g n_w = \frac{1.33}{1.5} = 0.89$.

From (i) above, it follows that in general

$${}_1 n_3 = {}_1 n_2 \times {}_2 n_3 \quad \dots \dots \dots (4)$$

The order of the suffixes enables this formula to be easily memorised.

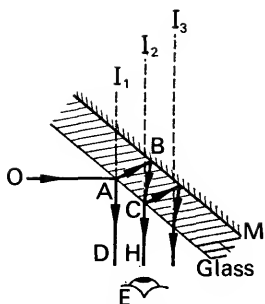


FIG. 18.5. Multiple images.

Suppose O is an object placed in front of a mirror with silvering on the back surface M , as shown in Fig. 18.5. A ray OA from O is then reflected from the front (glass) surface along AD and gives rise to a faint image I_1 , while the remainder of the light energy is refracted at A along AB . Reflection then takes place at the silvered surface, and after refraction into the air along CH a bright image is observed at I_2 . A small percentage of the light is reflected at C , however, and re-enters the glass again, thus forming a faint image at I_3 . Other faint images are formed in the same way. Thus a series of *multiple images* is obtained, the brightest being I_2 . The images lie on the normal from O to the mirror, the distances depending on the thickness of the glass and its refractive index and the angle of incidence.

Drawing the Refracted Ray by Geometrical Construction

Since $\sin i / \sin r = n$, the direction of the refracted ray can be calculated when a ray is incident in air at a known angle i on a medium of given refractive index n . The direction of the refracted ray can also be obtained by means of a geometrical construction. Thus suppose AO is a ray incident in air at a given angle i on a medium of refractive index n , Fig. 18.6 (i). With O as centre, two circles, a , b , are drawn whose radii are in the ratio $1:n$, and AO is produced to cut circle a at P . PN is then drawn parallel to the normal at O to intersect circle b at Q . OQ is then the direction of the refracted ray.

To prove the construction is correct, we note that angle $OPN = i$, angle $OQN = r$. Thus $\sin i / \sin r = ON/OP \div ON/OQ = OQ/OP$. But $OQ/OP = \text{radius of circle } b / \text{radius of circle } a = n$, from our drawing

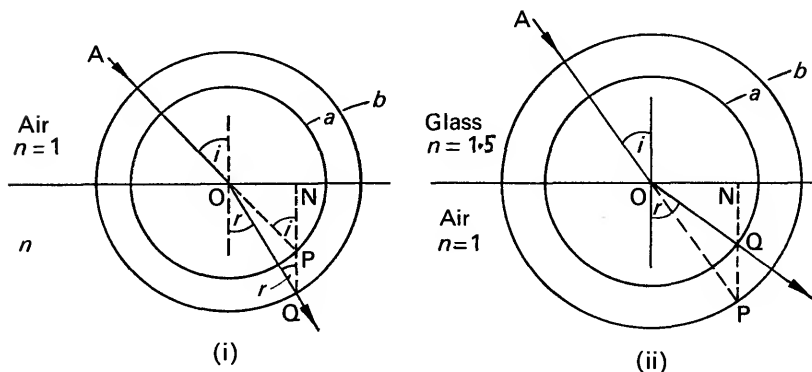


FIG. 18.6. Drawing of refracted rays.

of the circles. Hence $\sin i / \sin r = n$. Thus OQ must be the refracted ray. Although we have taken the case of the incident ray in air, the same construction will enable the refracted ray to be drawn when the incident ray is in any other medium. The radii of the circles are then in the ratio of the absolute refractive indices.

Fig. 18.6 (ii) illustrates the drawing in the case of a ray AO refracted from a dense medium such as glass ($n = 1.5$) into a less dense medium such as air ($n = 1$). Circles a , b are again drawn concentric with O, their radii being in the ratio $1 : n$. The incident ray AO, however, is produced to cut the *larger* circle b this time, at P, and a line PN is then drawn parallel to the normal at O to intersect the circle a at Q. OQ is then the direction of the refracted ray. The proof for the construction follows similar lines to that given for Fig. 18.6 (i), and it is left as an exercise for the reader.

The direction of the incident ray AO in Fig. 18.6 (ii) may be such that the line PN does not intersect the circle a . In this case, which is important and is discussed shortly, no refracted ray can be drawn. See *Total Internal Reflection*, p. 430.

Refractive Index of a Liquid by Using a Concave Mirror

We are now in a position to utilise our formulæ in refraction, and we shall first consider a simple method of determining roughly the refractive index, n , of a small quantity of transparent liquid.

If a small drop of the liquid is placed on a concave mirror S, a position H can be located by the no parallax method where the image of a pin held over the mirror coincides in position with the pin itself, Fig. 18.7. The rays from the pin must now be striking the mirror *normally*, in which case the rays are reflected back along the incident path and form an image at the same place as the object. A ray HN close to the axis HP is refracted at N along ND in the liquid, strikes the mirror normally at D, and is reflected back along the path DNH. Thus if DN is produced it passes through the centre of curvature, C, of the concave mirror.

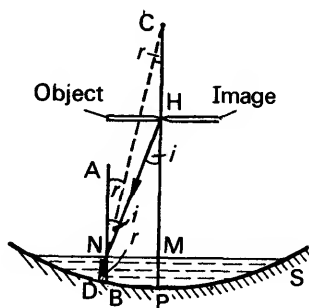


FIG. 18.7. (Depth of liquid exaggerated.)

Let ANB be the normal to the liquid surface at N. Then angle ANH = angle NHM = i , the angle of incidence, and angle BND = angle ANC = angle NCM = r , the angle of refraction in the liquid. From triangles HNM, CNM respectively, $\sin i = NM/HN$ and $\sin r = NM/CN$. The refractive index, n , of the liquid is thus given by

$$n = \frac{\sin i}{\sin r} = \frac{NM/HN}{NM/CN} = \frac{CN}{HN}.$$

Now since HN is a ray very close to the principal axis CP, $HN = HM$

and $CN = CM$, to a very good approximation. Thus $n = \frac{CM}{HM}$. Further,

if the depth MP of the liquid is very small compared with HM and CM, $CM = CP$ and $HM = HP$ approximately. Hence, approximately,

$$n = \frac{CP}{HP}.$$

HP can be measured directly. CP is the radius of curvature of the mirror, which can be obtained by the method shown on page 414. The refractive index, n , of the liquid can thus be calculated.

Apparent Depth

Swimmers in particular are aware that the bottom of a pool of water appears nearer the surface than is actually the case; the phenomenon is due to the refraction of light.

Consider an object O at a distance below the surface of a medium such as water or glass, which has a refractive index n , Fig. 18.8. A ray OM from O perpendicular to the surface passes straight through into the air along MS. A ray ON very close to OM is refracted at N into the air away from the normal, in a direction NT; and an observer viewing O directly overhead sees it in the position I, which is the point of intersection of SM and TN produced. Though we have only considered two rays in the air, a cone of rays, with SM as the axis, actually enters the observer's eye.

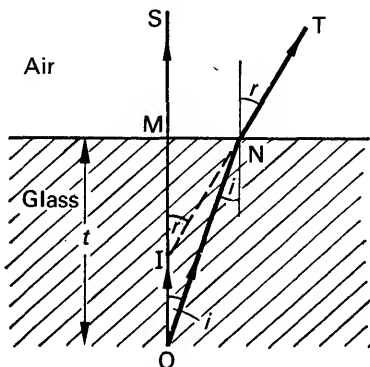


FIG. 18.8. (Inclination of ON to OM exaggerated.)

Suppose the angle of incidence in the glass is i , and the angle of refraction in the air is r . Then, since " $n \sin i$ " is a constant (p. 423), we have

$$n \sin i = 1 \times \sin r \quad \dots \dots \dots (i)$$

where n is the refractive index of glass; the refractive index of air is 1.

Since $i = \angle NOM$, and $r = \angle MIN$, $\sin i = MN/ON$ and $\sin r = MN/IN$. From (i), it follows that

$$n \frac{MN}{ON} = \frac{MN}{IN}$$

$$\therefore n = \frac{ON}{IN} \quad \dots \dots \dots (ii)$$

Since we are dealing with the case of an observer directly above O, the rays ON, IN are *very* close to the normal OM. Hence to a very good approximation, $ON = OM$ and $IN = IM$. From (ii),

$$\therefore n = \frac{ON}{IN} = \frac{OM}{IM}.$$

Since the real depth of the object O = OM, and its apparent depth = IM,

$$\therefore n = \frac{\text{real depth}}{\text{apparent depth}} \quad \dots \dots \dots (6)$$

If the real depth, OM, = t , the apparent depth = $\frac{t}{n}$, from (6). The displacement, OI, of the object, which we shall denote by d , is thus given by $t - \frac{t}{n}$, i.e.,

$$d = t \left(1 - \frac{1}{n} \right) \quad \dots \dots \dots (7)$$

If an object is 6 cm below water of refractive index, $n = 1\frac{1}{3}$, it appears to be displaced upward to an observer in air by an amount, d
 $= 6 \left(1 - \frac{1}{1\frac{1}{3}} \right) = 1\frac{1}{3}$ cm.

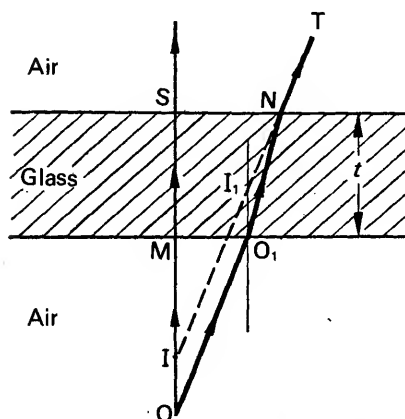


FIG. 18.9. Object below glass block.

Object Below Parallel-sided Glass Block

Consider an object O placed some distance in air below a parallel-sided glass block of thickness t , Fig. 18.9. The ray OMS normal to the surface emerges along MS, while the ray OO_1 close to the normal is refracted along O_1N in the glass and emerges in air along NT in a direction parallel to OO_1 (see p. 422). An observer (not shown) above the glass thus sees the object at I, the point of intersection of TN and SM.

Suppose the normal at O_1 intersects OI at I_1 . Then, since O_1I_1 is parallel to OI and IT is parallel to OO_1 , OII_1O_1 is a parallelogram. Thus $OI = O_1I_1$. But OI is the displacement of the object O. Hence O_1I_1 is

equal to the displacement. Since the apparent position of an object at O_1 is at I_1 (compare Fig. 18.8), we conclude that the displacement OI of O is independent of the position of O below the glass, and is given by

$$OI = t \left(1 - \frac{1}{n} \right), \text{ see p. 427.}$$

Measurement of Refractive Index by Apparent Depth Method

The formula for the refractive index of a medium in terms of the real and apparent depths is the basis of a very accurate method of measuring refractive index. A *travelling microscope*, S (a microscope which can travel in a vertical direction and which has a fixed graduated scale T beside it) is focused on lycopodium particles at O on a sheet of white paper, and the reading on T is noted, Fig. 18.10. Suppose it is c cm. If the refractive index of glass is required, a glass block A is placed on the paper, and the microscope is raised until the particles are refocused at I . Suppose the reading on T is b cm. Some lycopodium particles are then sprinkled at M on the top of the glass block, and the microscope is raised until they are focused, when the reading on T is noted. Suppose it is a cm.

Then real depth of $O = OM = (a - c)$ cm.

and apparent depth = $IM = (a - b)$ cm.

$$\therefore n = \frac{\text{real depth}}{\text{apparent depth}} = \frac{a - c}{a - b}$$

The high accuracy of this determination of n lies mainly in the fact that the objective of the microscope collects only those rays near to its axis, so that the object O , and its apparent position I , are seen by rays very close to the normal OM . The experiment thus fulfils the theoretical

conditions assumed in the proof of the formula $n = \frac{\text{real depth}}{\text{apparent depth}}$ p. 427.

The refractive index of water can also be obtained by an apparent depth method. The block A is replaced by a dish, and the microscope is focused first on the bottom of the dish and then on lycopodium powder sprinkled on the surface of water poured into the dish. The apparent position of the bottom of the dish is also noted, and the refractive index of the water n_w is calculated from the relation

$$n_w = \frac{\text{real depth of water}}{\text{apparent depth of water}}$$

General formula for real and apparent depth. So far we have considered the rays refracted from a medium like glass into air. As a more general case, suppose an object O is in a medium of refractive index n_1 and the rays from it are refracted at M , N into a medium of refractive index n_2 , Fig. 18.11. The image of O to an observer in the latter medium is then at I .

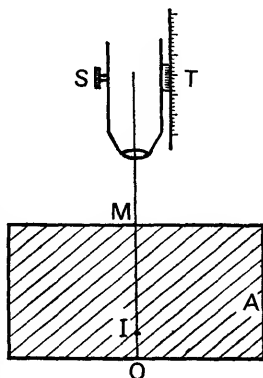


FIG. 18.10. Refractive index by apparent depth.

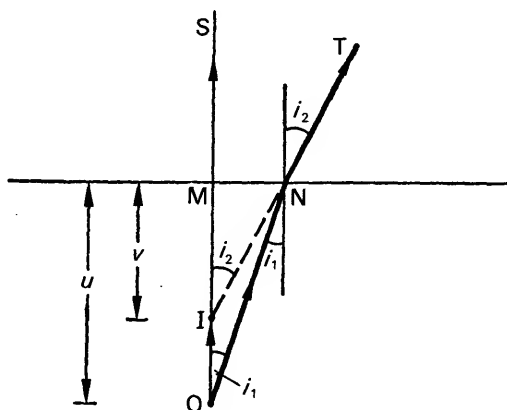


FIG. 18.11. Relation between real and apparent depth.

Suppose i_1 is the angle of incidence at N, and i_2 is the angle of refraction as shown. Then, since " $n \sin i$ " is a constant,

$$n_1 \sin i_1 = n_2 \sin i_2.$$

Now angle $MON = i_1$, and angle $MIN = i_2$.

$$\therefore n_1 \frac{MN}{ON} = n_2 \frac{MN}{IN}$$

$$\therefore \frac{n_1}{ON} = \frac{n_2}{IN} \quad \dots \dots \dots (i)$$

If we consider rays very close to the normal, then $IN = IM = v$, say, and $ON = OM = u$, say. Substituting in (i),

$$\therefore \frac{n_1}{u} = \frac{n_2}{v} \quad \dots \dots \dots (ii)$$

This formula can easily be remembered, as the refractive index of a medium is divided by the corresponding distance of the object (or image) *in* that medium.

Total Internal Reflection. Critical Angle

If a ray AO in glass is incident at a small angle α on a glass-air plane boundary, observation shows that part of the incident light is reflected along OE in the glass, while the remainder of the light is refracted away from the normal at an angle β into the air. The reflected ray OE is weak, but the refracted ray OL is bright, Fig. 18.12 (i). This means that most of the incident light energy is transmitted, and a little is reflected.

When the angle of incidence, α , in the glass is increased, the angle of emergence, β , is increased at the same time; and at some angle of incidence c in the glass the refracted ray OL travels along the glass-air boundary, making the angle of refraction of 90° , Fig. 18.12 (ii). The reflected ray OE is still weak in intensity, but as the angle of incidence in the glass is increased slightly the reflected ray suddenly becomes bright, and no refracted ray is then observed, Fig. 18.12 (iii). Since *all* the

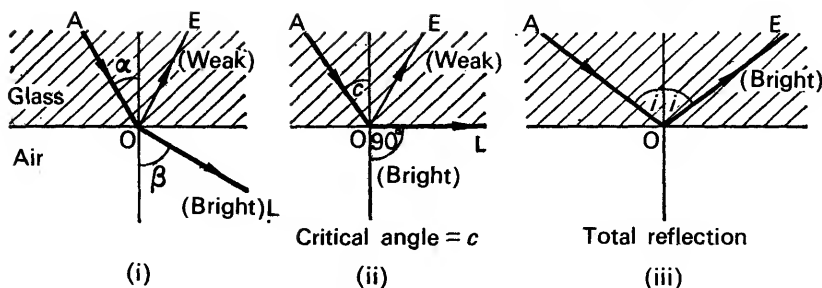


FIG. 18.12. Total internal reflection.

incident light energy is now reflected, **total reflection** is said to take place in the glass at O.

When the angle of refraction in air is 90° , a critical stage is reached at the point of incidence O, and the angle of incidence *in the glass* is accordingly known as the **critical angle** for glass and air, Fig. 18.12 (ii). Since " $n \sin i$ " is a constant (p. 423), we have

$$n \sin c = 1 \times \sin 90^\circ,$$

where n is the refractive index of the glass. As $\sin 90^\circ = 1$, then

$$n \sin c = 1,$$

or,

$$\sin c = \frac{1}{n} \quad . \quad . \quad . \quad (8)$$

Crown glass has a refractive index of about 1.51 for yellow light, and thus the critical angle for glass to air is given by $\sin c = 1/1.51 = 0.667$. Consequently $c = 41.5^\circ$. Thus if the incident angle in the glass is greater than c , for example 45° , total reflection occurs, Fig. 18.12 (iii). The critical angle between two media for blue light is less than for red light, since the refractive index for blue light is greater than that for red light (see p. 458).

The phenomenon of total reflection may occur when light in glass ($n_g = 1.51$, say) is incident on a boundary with water ($n_w = 1.33$). Applying " $n \sin i$ is a constant" to the critical case, Fig. 18.13, we have

$$n_g \sin c = n_w \sin 90^\circ,$$

where c is the critical angle. As $\sin 90^\circ = 1$

$$n_g \sin c = n_w$$

$$\therefore \sin c = \frac{n_w}{n_g} = \frac{1.33}{1.51} = 0.889$$

$$\therefore c = 63^\circ.$$

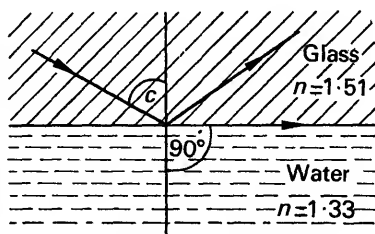


FIG. 18.13. Critical angle for water and glass.

Thus if the angle of incidence in the glass exceeds 63° , total internal reflection occurs.

It should be carefully noted that the phenomenon of total internal reflection can occur only when light travels from one medium to another which has a *smaller* refractive index, i.e., which is optically less dense. The phenomenon cannot occur when light travels from one medium to another optically denser, for example from air to glass, or from water to glass, as a refracted ray is then always obtained.

SOME APPLICATIONS OF TOTAL INTERNAL REFLECTION

1. *Reflecting prisms* are pieces of glass of a special shape which are used in prism binoculars and in certain accurate ranging instruments such as submarine periscopes. These prisms, discussed on p. 449, act as reflectors of light by total internal reflection.

2. *The mirage* is a phenomenon due to total reflection. In the desert the air is progressively hotter towards the sand, and hence the density of the air decreases in the direction bcd , Fig. 18.14 (i). A downward ray OA from a tree or the sky is thus refracted more and more away from the normal; but at some layer of air c , a critical angle is reached, and the ray begins to travel in an upward direction along cg . A distant observer P thus sees the object O at I , and hence an image of a palm tree, for example, is seen below the actual position of the tree. As an image of part of the sky is also formed by total reflection round the image of the tree, the whole appearance is similar to that of a pool of water in which the tree is reflected.

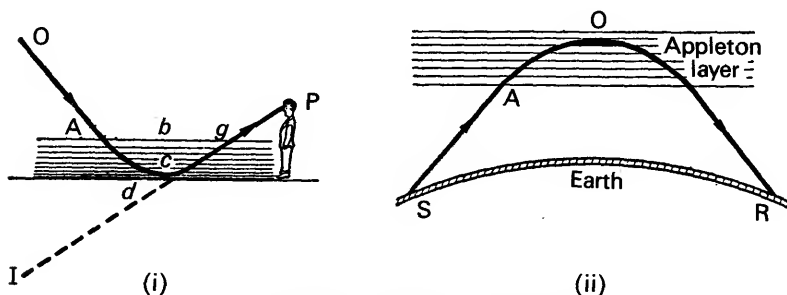


FIG. 18.14. Examples of total reflection.

3. *Total reflection of radio waves.* A radio wave is an example of an *electromagnetic wave* because it comprises electric and magnetic forces. Light waves also are electromagnetic waves (p. 983). Light waves and radio waves are therefore the same in nature, and a close analogy can be made between the refraction of light and the refraction of a radio wave when the latter enters a medium containing electric particles.

In particular, the phenomenon of total reflection occurs when radio waves travel from one place, S , on the earth, for example England, to another place, R , on the other side of the earth, for example America, Fig. 18.14 (ii). A layer of considerable density of electrons exists many

miles above the earth (at night this is the *Appleton layer*), and when a radio wave SA from a transmitter is sent skyward it is refracted away from the normal on entering the electron layer. At some height, corresponding to O, a critical angle is reached, and the wave then begins to be refracted downward. After emerging from the electron layer it returns to R on the earth, where its presence can be detected by a radio receiver.

Measurement of Refractive Index of a Liquid by an Air-cell Method

The phenomenon of total internal reflection is utilised in many methods of measuring refractive index. Fig. 18.15 (i) illustrates how the

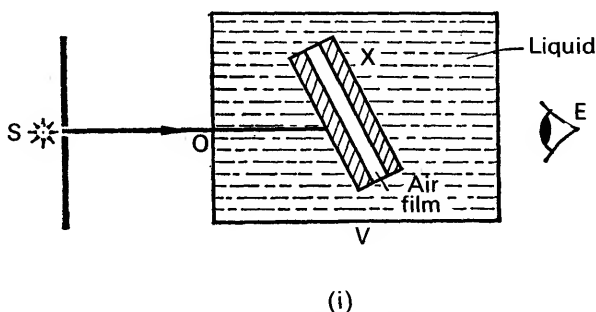


FIG. 18.15 (i). Air-cell method.

refractive index of a liquid can be determined. Two thin plane-parallel glass plates, such as microscope slides, are cemented together so as to contain a thin film of air of constant thickness between them, thus forming an air-cell, X. The liquid whose refractive index is required is placed in a glass vessel V having thin plane-parallel sides, and X is placed in the liquid. A bright source of light, S, provides rays which are incident on one side of X in a constant direction SO, and the light through X is observed by a person on the other side at E.

When the light is incident normally on the sides of X, the light passes straight through X to E. When X is rotated slightly about a vertical axis, light is still observed; but as X is rotated farther, the light is suddenly cut off from E, and hence no light now passes through X, Fig. 18.15 (i).

Fig. 18.15 (ii) illustrates the behaviour of the light when this happens. The ray SO is refracted along OB in the glass, but at B *total internal reflection begins*. Suppose i_1 is the angle of incidence in the liquid, i_2 is the angle of incidence in the glass, and n , n_g are the corresponding refractive indices. Since the boundaries of the media are parallel we can apply the relation " $n \sin i$ is a constant", and hence

$$n \sin i_1 = n_g \sin i_2 = 1 \times \sin 90^\circ, \quad \dots \quad (i)$$

the last product corresponding to the case of refraction in the air-film.

$$\therefore n \sin i_1 = 1 \times \sin 90^\circ = 1 \times 1 = 1$$

$$\therefore n = \frac{1}{\sin i_1} \quad \dots \quad (9)$$

It should now be carefully noted that i_1 is the angle of incidence in the *liquid* medium, and is thus determined by measuring the rotation of X from its position when normal to SO to the position when the light

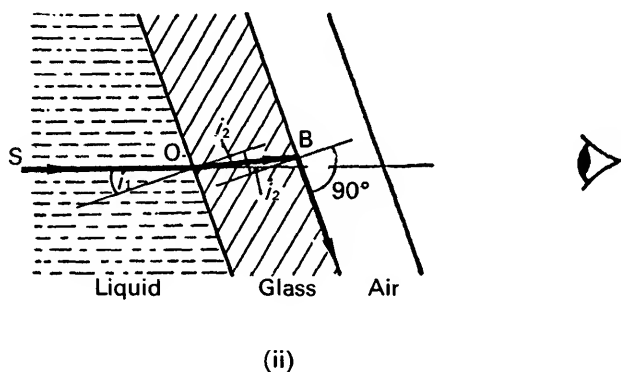


FIG. 18.15 (ii). Air-cell theory.

is cut off. In practice, it is better to rotate X in opposite directions and determine the angle θ between the *two* positions for the extinction of the light. The angle i_1 is then half the angle θ , and hence $n = 1/\sin \frac{\theta}{2}$.

From equations (i) and (9), it will be noted that i_1 is the critical angle between the liquid and air, and i_2 is the critical angle between the glass and air. We cannot measure i_2 , however, as we can i_1 , and hence the method provides the refractive index of the *liquid*.

The source of light, S, in the experiment should be a monochromatic source, i.e., it should provide light of one colour, for example yellow light. The extinction of the light is then sharp. If white light is used, the colours in its spectrum are cut off at slightly different angles of incidence, since refractive index depends upon the colour of the light (p. 458). The extinction of the light is then gradual and ill-defined.

Pulfrich Refractometer

A *refractometer* is an instrument which measures refractive index by making use of total internal reflection. PULFRICH designed a refractometer enabling the refractive index of a liquid to be easily obtained, which consists of a block of glass G with a polished and vertical face. On top of G is cemented a circular glass tube V, Fig. 18.16. The liquid L is placed in V, and a convergent beam of monochromatic light is directed so that the liquid-glass interface is illuminated. On observing the light refracted through G by a telescope T, a light and dark field of view are seen.

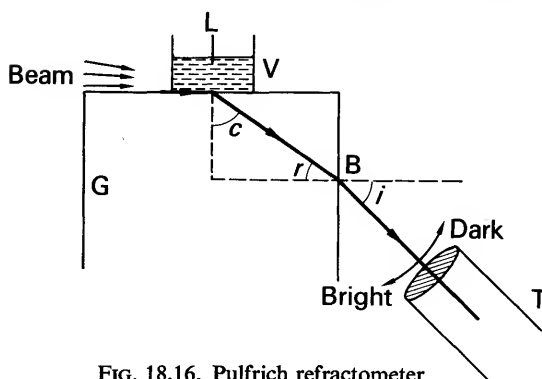


FIG. 18.16. Pulfrich refractometer.

The boundary between the light and dark fields corresponds to the ray which is incident just horizontally on the liquid-glass boundary, as shown in Fig. 18.16. If c is the angle of refraction in the glass, it follows that

$$n \sin 90^\circ = n_g \sin c \quad (i)$$

where n, n_g are the refractive indices of the liquid and glass respectively. For refraction at B,

$$n_g \sin r = \sin i \quad (ii)$$

$$\text{Also,} \quad c + r = 90^\circ \quad (iii)$$

From (i), $\sin c = n/n_g$. Now from (iii), $\sin r = \sin(90^\circ - c) = \cos c$. Substituting in (ii), we have $n_g \cos c = \sin i$, or $\cos c = \sin i/n_g$.

But $\sin^2 c + \cos^2 c = 1$

$$\therefore \frac{n^2}{n_g^2} + \frac{\sin^2 i}{n_g^2} = 1$$

$$\therefore n^2 + \sin^2 i = n_g^2$$

$$\therefore n = \sqrt{n_g^2 - \sin^2 i}$$

Thus if i is measured and $n_g = 1.51$, n can be calculated. In practice, tables are supplied giving the refractive index in terms of i , and another block is used in place of G for liquids of higher n than 1.51.

Abbe refractometer. Abbe designed a refractometer for measuring the refractive index of liquids whose principle is illustrated in Fig. 18.17. Two similar prisms X, Y are placed on a table A, the prism X being hinged at H so that it could be swung away from Y. A drop of the liquid is placed on the surface a , which is matt, and the prisms are placed together so that the liquid is squeezed into a thin film between them. Light from a suitable source is directed towards the prisms by means of a mirror M, where it strikes the surface a and is scattered by the matt surface into the liquid film. The emergent rays are collected in a telescope T directed towards the prisms, and the field of view is divided into a dark and bright portion. The table A is then turned until the dividing line between the dark and bright fields is on the crosswires of the telescope, which is fixed. The reading on the scale S, which is attached to, and moves with, the table, gives the refractive index of the liquid directly, as explained below.

EXAMPLES

1. Describe a method, based on grazing incidence or total internal reflection, for finding the refractive index of water for the yellow light emitted by a sodium flame.

The refractive index of carbon bisulphide for red light is 1.634 and the difference between the critical angles for red and blue light at a carbon bisulphide-air interface is $0^\circ 56'$. What is the refractive index of carbon bisulphide for blue light? (*N.*)

First part. See air-cell method, p. 432.

Second part. Suppose n_r and c_r are the refractive index and critical angle of carbon bisulphide for red light.

$$\text{Then} \quad \sin c_r = \frac{1}{n_r} = \frac{1}{1.634} = 0.6119$$

$$\therefore c_r = 37^\circ 44'$$

The critical angle, c_b , for blue light is less than that for red light.

$$\therefore c_b = 37^\circ 44' - 0^\circ 56' = 36^\circ 48'$$

$$\text{The refractive index for blue light, } n_b = \frac{1}{\sin c_b}$$

$$\therefore n_b = \frac{1}{\sin 36^\circ 48'} = 1.669$$

2. Find an expression for the distance through which an object appears to be displaced towards the eye when a plate of glass of thickness t and refractive index n is interposed.

A tank contains a slab of glass 8 cm thick and of refractive index 1.6. Above this is a depth of 4.5 cm of a liquid of refractive index 1.5 and upon this floats 6 cm of water ($n = 4/3$). To an observer looking from above, what is the apparent position of a mark on the bottom of the tank? (*O.* & *C.*)

First part. See text.

Second part. Suppose *O* is the mark at the bottom of the tank, Fig. 18.19. Then

since the boundaries of the media are parallel, the total displacement of *O* is the sum of the displacements due to each of the media.

$$\text{For glass, displacement, } d = t \left(1 - \frac{1}{n} \right) = 8 \left(1 - \frac{1}{1.6} \right) = 3 \text{ cm}$$

$$\text{For liquid, } d = t \left(1 - \frac{1}{n} \right) = 4.5 \left(1 - \frac{1}{1.5} \right) = 1.5 \text{ cm}$$

$$\text{For water, } d = 6 \left(1 - \frac{1}{4/3} \right) = 1.5 \text{ cm}$$

$$\therefore \text{total displacement} = 3 + 1.5 + 1.5 = 6 \text{ cm}$$

$$\therefore \text{apparent position of } O \text{ is 6 cm from bottom.}$$

3. A small object is placed on the principal axis of a concave spherical mirror of radius 20 cm at a distance of 30 cm. By how much will the position

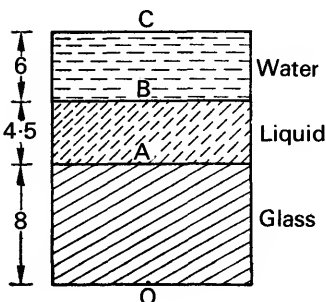


FIG. 18.19.

and size of the image alter when a parallel-sided slab of glass, of thickness 6 cm and refractive index 1.5, is introduced between the centre of curvature and the object? The parallel sides are perpendicular to the principal axis. Prove any formula used. (N.)

Suppose O is the position of the object before the glass is placed in position,

Fig. 18.20. The image position is given by $\frac{1}{v} + \frac{1}{u} = \frac{2}{r}$,

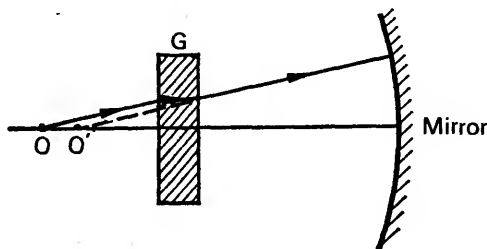


FIG. 18.20.

$$\therefore \frac{1}{v} + \frac{1}{30} = \frac{2}{20}$$

Solving, $v = 15$ cm.

The magnification, $m = \frac{v}{u} = \frac{15}{30} = 0.5$.

When the glass slab G of thickness, t , 6 cm is inserted, the rays from O appear to come from a point O' whose displacement from O is $t \left(1 - \frac{1}{n}\right)$, where n is the glass refractive index. See p. 428. The displacement is thus $6 \left(1 - \frac{1}{1.5}\right) = 2$ cm. The distance of O' from the mirror is therefore $(30 - 2)$, or 28 cm. Applying the equation $\frac{1}{v} + \frac{1}{u} = \frac{2}{r}$, we find $v = +15\frac{5}{8}$ cm. The image position changes by $(15\frac{5}{8} - 15)$ or $\frac{5}{8}$ cm. The magnification becomes $15\frac{5}{8} \div 28$, or 0.52.

EXERCISES 18

1. A ray of light is incident at 60° in air on an air-glass plane surface. Find the angle of refraction in the glass by calculation and by drawing (n for glass = 1.5).

2. A ray of light is incident in water at an angle of 30° on a water-air plane surface. Find the angle of refraction in the air by calculation and by drawing (n for water = $4/3$).

3. A ray of light is incident in water at an angle of (i) 30° , (ii) 70° on a water-glass plane surface. Calculate the angle of refraction in the glass in each case ($n_g = 1.5$, $n_w = 1.33$).

4. (a) Describe the apparent depth method of finding the refractive index of glass, and prove the formula used. (b) What is the apparent position of an object below a rectangular block of glass 6 cm thick if a layer of water 4 cm thick is on top of the glass (refractive index of glass and water = $1\frac{1}{2}$ and $1\frac{1}{3}$ respectively)?

5. Describe and explain a method of measuring approximately the refractive index of a small quantity of liquid.

6. Calculate the critical angle for (i) an air-glass surface, (ii) an air-water surface, (iii) a water-glass surface; draw diagrams in each case illustrating the total reflection of a ray incident on the surface ($n_g = 1.5$, $n_w = 1.33$).

7. Explain what happens in general when a ray of light strikes the surface separating transparent media such as water and glass. Explain the circumstances in which total reflection occurs and show how the critical angle is related to the refractive index.

Describe a method for determining the refractive index of a medium by means of critical reflection. (L.)

8. Define *refractive index* of one medium with respect to another and show how it is related to the values of the velocity of light in the two media.

Describe a method of finding the refractive index of water for sodium light, deducing any formula required in the reduction of the observations. (N.)

9. Explain carefully why the apparent depth of the water in a tank changes with the position of the observer.

A microscope is focused on a scratch on the bottom of a beaker. Turpentine is poured into the beaker to a depth of 4 cm, and it is found necessary to raise the microscope through a vertical distance of 1.28 cm to bring the scratch again into focus. Find the refractive index of the turpentine. (C.)

10. What is meant by *total reflection* and *critical angle*? Describe two methods of measuring the refractive index of a material by determining the critical angle, one of which is suitable for a solid substance and the other for a liquid. (L.)

11. (a) State the conditions under which total reflection occurs. Show that the phenomenon will occur in the case of light entering normally one face of an isosceles right-angle prism of glass, but not in the case when light enters similarly a similar hollow prism full of water. (b) A concave mirror of small aperture and focal length 8 cm lies on a bench and a pin is moved vertically above it. At what point will image and object coincide if the mirror is filled with water of refractive index $4/3$? (N.)

12. State the laws of refraction, and define *refractive index*.

Describe an accurate method of determining the refractive index of a transparent liquid for sodium light. Give the theory of the method, and derive any formula you require. Discuss the effect of substituting white light for sodium light in your experiment. (W.)

13. Describe an experiment for finding the refractive index of a liquid by measuring its apparent depth.

A vessel of depth $2d$ cm is half filled with a liquid of refractive index n_1 , and the upper half is occupied by a liquid of refractive index n_2 . Show that

the apparent depth of the vessel, viewed perpendicularly, is $d \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$. (L.)

14. The base of a cube of glass of refractive index n_1 is in contact with the surface of a liquid of refractive index n_2 . Light incident on one vertical face of the cube is reflected internally from the base and emerges again from the opposite vertical face in a direction making an angle θ with its normal. Assuming that $n_1 > n_2$, show that the light has just been totally reflected internally if $n_2 = \sqrt{(n_1^2 - \sin^2 \theta)}$.

Describe how the above principle may be used to measure the refractive index of a small quantity of liquid. (N.)

15. Explain the meaning of critical angle and total internal reflection. Describe fully (a) one natural phenomenon due to total internal reflection, (b) one practical application of it. Light from a luminous point on the lower face of a rectangular glass slab, 2.0 cm thick, strikes the upper face and the totally reflected rays outline a circle of 3.2 cm radius on the lower face. What is the refractive index of the glass? (N.)

16. Describe one experiment in each case to determine the refractive index for sodium light of (a) a sample of glass which could be supplied to any shape and size which you specify, (b) a liquid of which only a very small quantity is available. Show how the result is calculated in each case. (You are *not* expected to derive standard formulae.)

How would you modify the experiment in (a) to find how the refractive index varies with the wavelength of the light used? What general result would you expect? (O. & C.)

17. Explain the meaning of *critical angle*, and describe how you would measure the critical angle for a water-air boundary.

ABCD is the plan of a glass cube. A horizontal beam of light enters the face AB at grazing incidence. Show that the angle θ which any rays emerging from BC would make with the normal to BC is given by $\sin \theta = \cot a$, where a is the critical angle. What is the greatest value that the refractive index of glass may have if any of the light is to emerge from BC? (N.)

18. State the *laws of refraction of light*. Explain how you would measure the refractive index of a transparent liquid available only in *small* quantity, i.e., less than 0.5 cm³.

A ray of light is refracted through a sphere, whose material has a refractive index n , in such a way that it passes through the extremities of two radii which make an angle a with each other. Prove that if γ is the deviation of the ray caused by its passage through the sphere

$$\cos \frac{1}{2} (a - \gamma) = n \cos \frac{1}{2} a. \quad (L.)$$

19. Explain what is meant by the terms *critical angle* and *total reflection*. Describe an accurate method of determining the critical angle for a liquid, indicating how you would calculate the refractive index from your measurements.

A man stands at the edge of the deep end of a swimming bath, the floor of which is covered with square tiles. If the water is clear and undisturbed, explain carefully how the floor of the bath appears to him. (O. & C.)

20. Summarize the various effects that may occur when a parallel beam of light strikes a plane interface between two transparent media.

Explain why, when looking at the windows of a railway carriage from inside, one sees by day the country outside and by night the reflection of the inside of the carriage.

An observer looks normally through a thick window of thickness d and refractive index n at an object at a distance e behind the farther surface. Where does the object appear to be, and how can this apparent position be found experimentally? (*O. & C.*)

Refraction through prisms

441

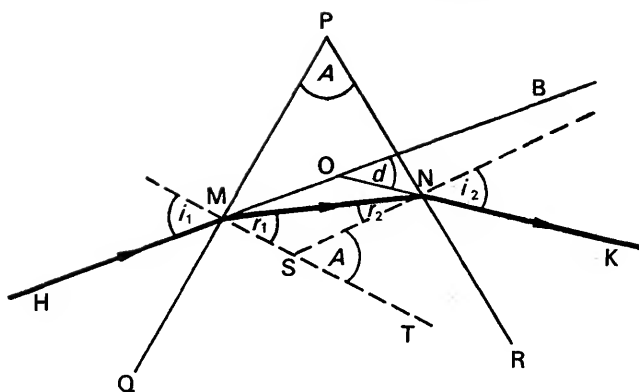


FIG. 19.2. Refraction through prism.

Further, as MS and NS are normals to PM and PN respectively, angle MPN + angle MSN = 180° , considering the quadrilateral PMSN. But angle NST + angle MSN = 180° .

$$\therefore \text{angle NST} = \text{angle MPN} = A$$

$$\therefore A = r_1 + r_2 \quad \dots \quad (iii)$$

as angle NST is the exterior angle of triangle MSN.

In the following sections, we shall see that the *angle of deviation*, d , of the light, caused by the prism, is utilised considerably. The angle of deviation at M = angle OMN = $i_1 - r_1$; the angle of deviation at N = angle MNO = $i_2 - r_2$. Since the deviations at M, N are in the same direction, the total deviation, d (angle BOK), is given by

$$d = (i_1 - r_1) + (i_2 - r_2) \quad \dots \quad (iv)$$

Equations (i) – (iv) are the general relations which hold for refraction through a prism. In deriving them, it should be noted that the geometrical form of the prism base plays no part.

Minimum Deviation

The angle of deviation, d , of the incident ray HM is the angle BOK in Fig. 19.2. The variation of d with the angle of incidence, i , can be obtained experimentally by placing the prism on paper on a drawing board and using a ray AO from a ray-box (or two pins) as the incident ray, Fig. 19.3 (i). When the direction AO is kept constant and the drawing board is turned so that the ray is always incident at O on the prism, the angle of incidence i is varied; the corresponding emergent rays CE, HK, LM, NP can be traced on the paper. Experiment shows that as the angle of incidence i is increased from zero, the deviation d begins to decrease continuously to some value D , and then increases to a maximum as i is increased further to 90° . A *minimum deviation*, corresponding to the emergent ray NP, is thus obtained. A graph of d plotted against i has the appearance of the curve X, which has a minimum value at R, Fig. 19.3 (ii).

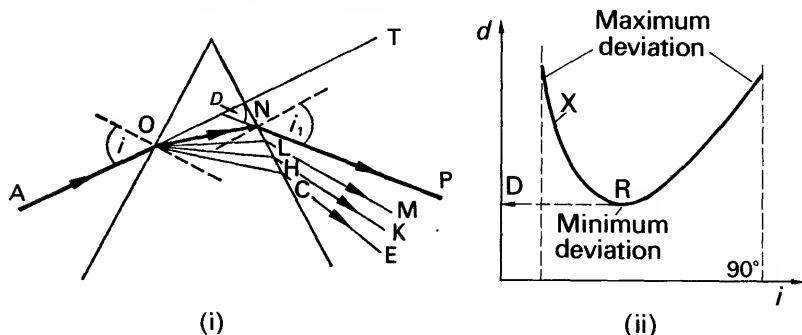


FIG. 19.3. Minimum deviation.

Experiment and theory show that *the minimum deviation, D , of the light occurs when the ray passes symmetrically through the prism*. Suppose this corresponds to the case of the ray AONP in Fig. 19.3 (i). Then the corresponding incident angle, i , is equal to the angle of emergence, i_1 , into the air at N for this special case. See also Fig. 19.5 and Fig. 19.9.

A proof of symmetrical passage of ray at minimum deviation. Experiment shows that minimum deviation is obtained at *one* particular angle of incidence. On this assumption it is possible to prove by a *reductio ad absurdum* method that the angle of incidence is equal to the angle of emergence in this case. Thus

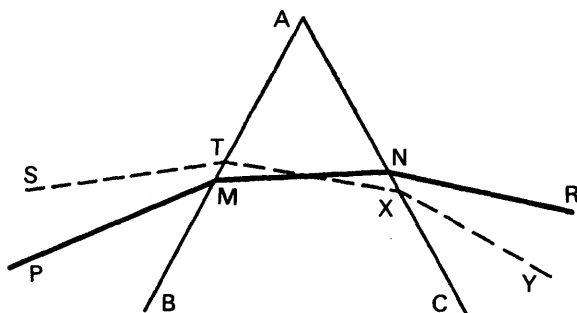


FIG. 19.4. Minimum deviation proof.

suppose that minimum deviation is obtained with a ray PMNR when these angles are *not* equal, so that angle PMB is not equal to angle RNC, Fig. 19.4. It then follows that a ray YX, incident on AC at an angle CXY equal to angle PMB, will emerge along TS, where angle BTS = angle CNR; and from the principle of the reversibility of light, a ray incident along ST on the prism emerges along XY. We therefore have *two* cases of minimum deviation, corresponding to two different angles of incidence. But, from experiment, this is impossible. Consequently our initial assumption must be wrong, and hence the angle of emergence *does* equal the angle of incidence. Thus the ray passes symmetrically through the prism in the minimum deviation case.

Relation Between A , D , and n

A very convenient formula for refractive index, n , can be obtained in the minimum deviation case. The ray PQRS then passes symmetrically through the prism, and the angles made with the normal in the air and in the glass at Q, R respectively are equal, Fig. 19.5. Suppose the angles are denoted by i , r , as shown. Then, as explained on p. 442,

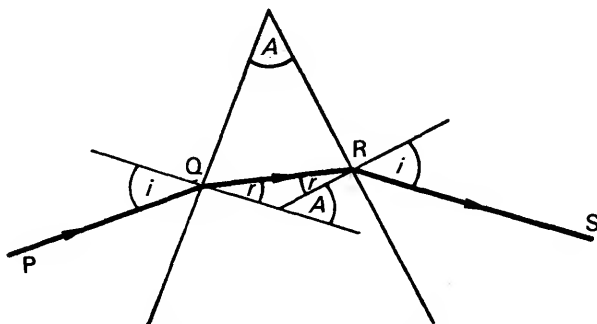


FIG. 19.5. Formula for n of prism.

$$i - r + i - r = D \quad . \quad . \quad . \quad . \quad . \quad (i)$$

$$\text{and} \quad r + r = A \quad . \quad . \quad . \quad . \quad . \quad (ii)$$

$$\text{From (ii),} \quad r = \frac{A}{2}$$

$$\text{Substituting for } r \text{ in (i),} \quad 2i = A + D$$

$$\therefore i = \frac{A + D}{2}$$

$$\therefore n = \frac{\sin i}{\sin r} = \frac{\sin \frac{A + D}{2}}{\sin \frac{A}{2}} \quad . \quad . \quad (1)$$

The Spectrometer

The spectrometer is an optical instrument which is mainly used to study the light from different sources. As we shall see later, it can be used to measure accurately the refractive index of glass in the form of a prism. The instrument consists essentially of a *collimator*, C, a *telescope*, T, and a *table*, R, on which a prism B can be placed. The lenses in C, T are achromatic lenses (p. 515). The collimator is fixed, but the table and the telescope can be rotated round a circular scale graduated in half-degrees (not shown) which has a common vertical axis with the table, Fig. 19.6. A vernier is also provided for this scale. The *source of light*, S, used in the experiment is placed in front of a narrow slit at one end of the collimator, so that the prism is illuminated by light from S.

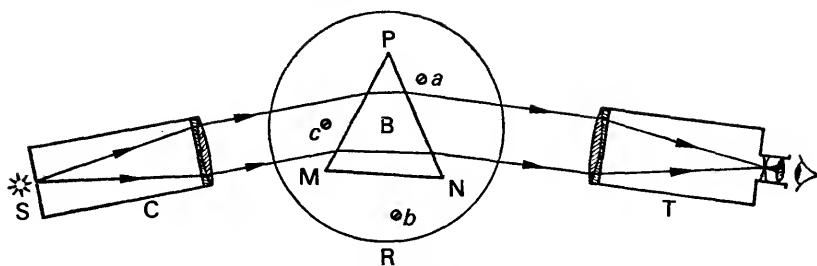


FIG. 19.6. Spectrometer.

Before the spectrometer can be used, however, three adjustments must be made: (1) The collimator *C* must be adjusted so that parallel light emerges from it; (2) the telescope *T* must be adjusted so that parallel rays entering it are brought to a focus at cross-wires near its eye-piece; (3) the refracting edge of the prism must be parallel to the axis of rotation of the telescope, i.e., the table must be “levelled”.

Adjustments of Spectrometer

The telescope adjustment is made by first moving its eye-piece until the cross-wires are distinctly seen, and then sighting the telescope on to a distant object through an open window. The length of the telescope is now altered by a screw arrangement until the object is clearly seen at the same place as the cross-wires, so that parallel rays now entering the telescope are brought to a focus at the cross-wires.

The collimator adjustment. With the prism removed from the table, the telescope is now turned to face the collimator, *C*, and the slit in *C* is illuminated by a sodium flame which provides yellow light. The edges of the slit are usually blurred, showing that the light emerging from the lens of *C* is not a parallel beam. The position of the slit is now adjusted by moving the tube in *C*, to which the slit is attached, until the edges of the latter are sharp.

“Levelling” the table. If the rectangular slit is not in the centre of the field of view when the prism is placed on the table, the refracting edge of the prism is not parallel to the axis of rotation of the telescope. The table must then be adjusted, or “levelled”, by means of the screws *a*, *b*, *c* beneath it. One method of procedure consists of placing the prism on the table with one face *MN* approximately perpendicular to the line joining two screws *a*, *b*, as shown in Fig. 19.6. The table is turned until *MN* is illuminated by the light from *C*, and the telescope *T* is then moved to receive the light reflected from *MN*. The screw *b* is then adjusted until the slit appears in the centre of the field of view. With *C* and *T* fixed, the table is now rotated until the slit is seen by reflection at the face *NP* of the prism, and the screw *c* is then adjusted until the slit is again in the middle of the field of view. The screw *c* moves *MN* in its own plane, and hence the movement of *c* will not upset the adjustment of *MN* in the perpendicular plane.

Measurement of the Angle, A , of a Prism

The angle of a prism can be measured very accurately by a spectrometer. The refracting edge, P , of the prism is turned so as to face the collimator lens, which then illuminates the two surfaces containing

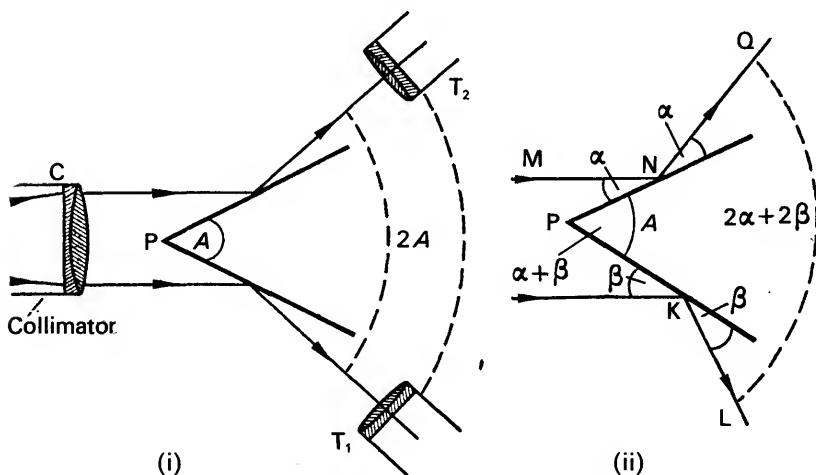


FIG. 19.7. Measurement of angle of prism.

the refracting angle A with parallel light, Fig. 19.7 (i). An image of the collimator slit is hence observed with the telescope in positions T_1 , T_2 , corresponding to reflection of light at the respective faces of the prism. It is shown below that the angle of rotation of the telescope from T_1 to T_2 is equal to $2A$, and hence the angle of the prism, A , can be obtained.

Proof. Suppose the incident ray MN makes a glancing angle α with one face of the prism, and a parallel ray at K makes a glancing angle β with the other face, Fig. 19.7 (ii). The reflected ray NQ then makes a glancing angle α with the prism surface, and hence the deviation of MN is 2α (see p. 392). Similarly, the deviation by reflection at K is 2β . Thus the reflected rays QN , LK are inclined at an angle equal to $2\alpha + 2\beta$, corresponding to the angle of rotation of the telescope from T_1 to T_2 . But the angle, A , of the prism $= \alpha + \beta$, as can be seen by drawing a line through P parallel to MN , and using alternate angles. Hence the rotation of the telescope $= 2\alpha + 2\beta = 2A$.

Measurement of the Minimum Deviation, D

In order to measure the minimum deviation, D , caused by refraction through the prism, the latter is placed with its refracting angle A pointing away from the collimator, as shown in Fig. 19.8 (i). The telescope is then turned until an image of the slit is obtained on the cross-wires, corresponding to the position T_1 . The table is now slowly rotated so that the angle of incidence on the left side of the prism decreases, and the image of the slit is kept on the cross-wires by moving the telescope at the same time. The image of the slit, and the telescope, then slowly approach the

fixed line XY. But at one position, corresponding to T_2 , the image of the slit begins to move away from XY. If the table is now turned in the opposite direction the image of the slit again moves back when the telescope reaches the position T_2 . The angle between the emergent ray CH and the line XY is hence the smallest angle of deviation caused by the prism, and is thus equal to D .

The minimum deviation is obtained by finding the angle between the positions of the telescope (i) at T_2 , (ii) at T; the prism is removed in the latter case so as to view the slit directly. Alternatively, the experiment to find the minimum deviation is repeated with the refracting angle pointing

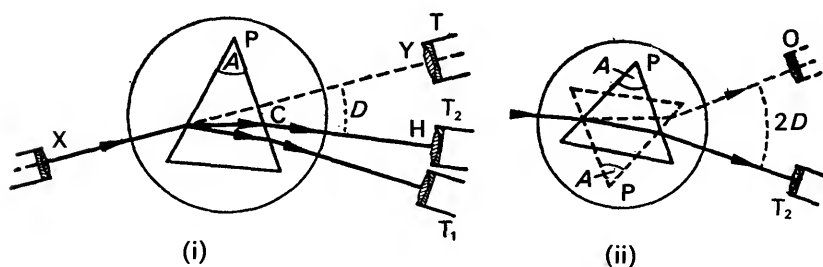


FIG. 19.8. Measurement of minimum deviation.

the opposite way, the prism being represented by dotted lines in this case, Fig. 19.8 (ii). If the position of the telescope for minimum deviation is now O, it can be seen that the angle between the position O and the other minimum deviation position T_2 is $2D$. The value of D is thus easily calculated.

The Refractive Index of the Prism Material

The refractive index, n , of the material of the prism can be easily calculated once A and D have been determined, since, from p. 444.

$$n = \sin \frac{A + D}{2} \bigg/ \sin \frac{A}{2}.$$

In an experiment of this nature, the angle, A , of a glass prism was found to be $59^\circ 52'$, and the minimum deviation, D , was $40^\circ 30'$. Thus

$$\begin{aligned} n &= \sin \frac{59^\circ 52' + 40^\circ 30'}{2} \bigg/ \sin \frac{59^\circ 52'}{2} \\ &= \sin 50^\circ 11' / \sin 29^\circ 56' \\ &= 1.539 \end{aligned}$$

The spectrometer prism method of measuring refractive index is capable of providing an accuracy of one part in a thousand. The refractive index of a liquid can also be found by this method, using a hollow glass prism made from thin parallel-sided glass strips.

ray MN in the glass will make a bigger and bigger angle of incidence on the other face PR, Fig. 19.10. This is left as an exercise for the reader. At

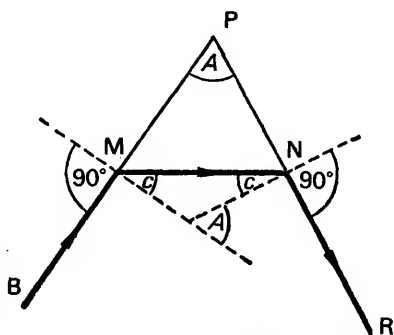


FIG. 19.10. Maximum angle of prism.

a certain value of A , MN will make the critical angle, c , with the normal at N, and the emergent ray NR will then graze the surface PR, as shown in Fig. 19.10. As A is increased further, the rays in the glass strike PR at angles of incidence greater than c , and hence no emergent rays are obtained. Thus Fig. 19.10 illustrates the largest angle of a prism for which emergent rays are obtained, and this is known as the *limiting angle* of the prism. It can be seen from the simple geometry of Fig. 19.10 that $A = c + c$ in this special case, and hence *the limiting angle of a prism is twice the critical angle*. For crown glass of $n = 1.51$ the critical angle c is $41^\circ 30'$, and hence transmission of light through a prism of crown glass is impossible if the angle of the prism exceeds 83° .

Total Reflecting Prisms

When a plane mirror silvered on the back is used as a reflector, multiple images are obtained (p. 424). This disadvantage is overcome by using right-angled isosceles prisms as reflectors of light in optical instruments such as submarine periscopes (see p. 540).

Consider a ray OQ incident normally on the face AC of such a prism. Fig. 19.11 (i). The ray is undeviated, and is therefore incident at P in the

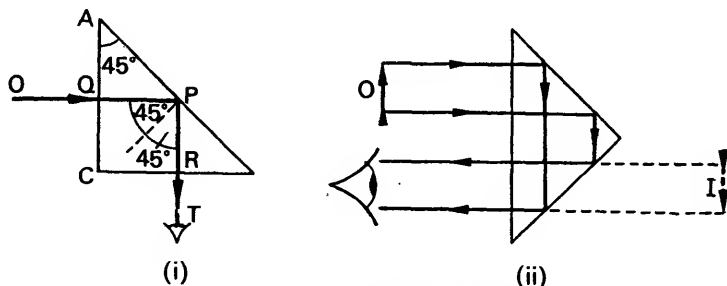


FIG. 19.11. Images in prisms.

glass at an angle of 45° to the normal at P. If the prism is made of crown glass its critical angle is $41^\circ 30'$. Hence the incident angle, 45° , in the glass is greater than the critical angle, and consequently the light is *totally* reflected in the glass at P. A bright beam of light thus emerges from the prism along RT, and since the angle of reflection at P is equal to the incident angle, RT is perpendicular to OQ. The prism thus deviates the light through 90° . If the prism is positioned as shown in Fig. 19.11 (ii), an inverted bright virtual image I of the object O is seen by total reflection at the two surfaces of the prism.

There is no loss of brightness when total internal reflection occurs at a surface, whereas the loss may be as much as 10 per cent or more in reflection at a silver surface.

EXAMPLES

1. Describe a good method of measuring the refractive index of a substance such as glass and give the theory of the method. A glass prism of angle 72° and index of refraction 1.66 is immersed in a liquid of refractive index 1.33. What is the angle of minimum deviation for a parallel beam of light passing through the prism? (L.)

First part. Spectrometer can be used, p. 444.

Second part.

$$n = \frac{\sin \left(\frac{A + D}{2} \right)}{\sin \frac{A}{2}}$$

where n is the *relative refractive index* of glass with respect to the liquid.

$$\text{But } n = \frac{1.66}{1.33}$$

$$\therefore \frac{1.66}{1.33} = \frac{\sin \left(\frac{72^\circ + D}{2} \right)}{\sin \frac{72^\circ}{2}} = \frac{\sin \left(\frac{72^\circ + D}{2} \right)}{\sin 36^\circ}$$

$$\therefore \sin \left(\frac{72^\circ + D}{2} \right) = \frac{1.66}{1.33} \sin 36^\circ = 0.7335$$

$$\therefore \frac{72^\circ + D}{2} = 47^\circ 11'$$

$$\therefore D = 22^\circ 22'$$

2. How would you measure the angle of minimum deviation of a prism?

(a) Show that the ray of light which enters the first face of a prism at grazing incidence is least likely to suffer total internal reflection at the other face.

(b) Find the least value of the refracting angle of a prism made of glass of refractive index $7/4$ so that no rays incident on one of the faces containing this angle can emerge from the other. (N.)

First part. See text.

Second part. (a) Suppose PM is a ray which enters the first face of the prism at grazing incidence, i.e., at an angle of incidence of 90° , Fig. 19.12. The refracted ray MQ then makes an angle of refraction c , where c is the critical angle. Suppose QN is the normal at Q on the other face of the prism. Then since angle $BNQ = A$, where A is the angle of the prism, the angle of refraction MQN at $Q = A - c$, from the exterior angle property of triangle MQN . Similarly, if RM is a ray at an angle of incidence i at the first face less than 90° , the angle of refraction at S at the second face $= A - r$, where r is the angle of refraction BMS .

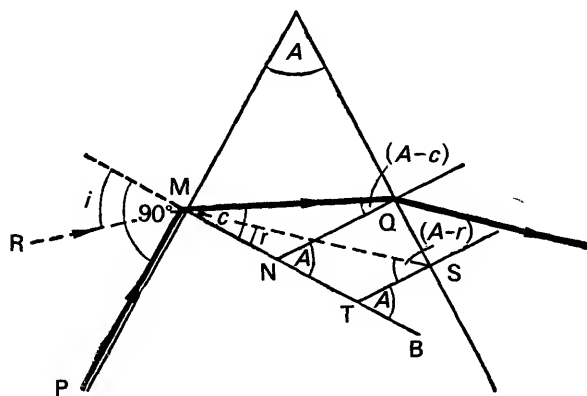


FIG. 19.12. Example

Now c is the *maximum* angle of refraction in the prism.

Hence MQ makes the minimum angle of incidence on the second face, and thus is least likely to suffer total internal reflection.

(b) The least value of the refracting angle of the prism corresponds to a ray at grazing incidence and grazing emergence, as shown in Fig. 19.10.

Thus minimum angle $= 2c$
where c is the critical angle (p. 449).

But

$$\sin c = \frac{1}{n} = \frac{4}{7} = 0.5714$$

$$\therefore c = 34^\circ 51'$$

$$\therefore \text{minimum angle} = 2c = 69^\circ 42'$$

EXERCISES 19

1. A ray of light is refracted through a prism of angle 70° . If the angle of refraction in the glass at the first face is 28° , what is the angle of incidence in the glass at the second face?

2. (i) The angle of a glass prism is 60° , and the minimum deviation of light through the prism is 39° . Calculate the refractive index of the glass.
(ii) The refractive index of a glass prism is 1.66, and the angle of the prism is 60° . Find the minimum deviation.

3. By means of a labelled diagram show the paths of rays from a monochromatic source to the eye through a correctly adjusted prism spectrometer.

Obtain an expression relating the deviation of the beam by the prism to the refracting angle and the angles of incidence and emergence.

A certain prism is found to produce a minimum deviation of $51^\circ 0'$, while it produces a deviation of $62^\circ 48'$ for two values of the angle of incidence, namely $40^\circ 6'$ and $82^\circ 42'$ respectively. Determine the refracting angle of the prism, the angle of incidence at minimum deviation and the refractive index of the material of the prism. (L.)

4. A ray of light passing symmetrically through a glass prism of refracting angle A is deviated through an angle D . Derive an expression for the refractive index of the glass.

A prism of refracting angle about 60° is mounted on a spectrometer table and all the preliminary adjustments are made to the instrument. Describe and explain how you would then proceed to measure the angles A and D .

PQR represents a right-angled isosceles prism of glass of refractive index 1.50. A ray of light enters the prism through the hypotenuse QR at an angle of incidence i , and is reflected at the critical angle from PQ to PR . Calculate and draw a diagram showing the path of the ray through the prism. (Only rays in the plane of PQR need be considered.) (N.)

5. Give a labelled diagram showing the essential optical parts of a prism spectrometer. Describe the method of adjusting a spectrometer and using it to measure the angle of a prism.

A is the vertex of a triangular glass prism, the angle at A being 30° . A ray of light OP is incident at P on one of the faces enclosing the angle A , in a direction such that the angle $OPA = 40^\circ$. Show that, if the refractive index of the glass is 1.50, the ray cannot emerge from the second face. (L.)

6. Define refractive index and derive an expression relating the relative refractive index n_{AB} for light travelling out of medium A into medium B with the velocities of light v_A and v_B respectively in those media.

Draw a diagram showing how a parallel beam of monochromatic light is deviated by its passage through a triangular glass prism. Given that the angle of deviation is a minimum when the angles of incidence and emergence are equal show that the refractive index n of the glass is related to the refracting angle α of the prism and the minimum deviation δ by the equation

$$n = \sin \frac{1}{2}(\alpha + \delta) / \sin \frac{1}{2}\alpha.$$

Describe how you would apply this result to measure the dispersive power of the glass of a given triangular prism. You may assume the availability of sources of light of standard wavelengths. (O. & C.)

7. Explain how you would adjust the telescope of a spectrometer before making measurements.

Draw and label a diagram of the optical parts of a prism spectrometer after the adjustments have been completed. Indicate the position of the crosswires and show the paths through the instrument of two rays from a monochromatic source when the setting for minimum deviation has been obtained.

The refracting angle of a prism is 62.0° and the refractive index of the glass for yellow light is 1.65. What is the smallest possible angle of incidence of a ray of this yellow light which is transmitted without total internal reflection? Explain what happens if white light is used instead, and the angle of incidence is varied in the neighbourhood of this minimum. (N.)

8. Explain the meaning of the term *critical angle*. Describe and give the theory of a critical angle method for determining the refractive index of water.

A right-angled prism ABC has angle $BAC = \text{angle } ACB = 45^\circ$, and is made of glass of refractive index 1.60. A ray of light is incident upon the hypotenuse face AC so that after refraction it strikes face AB and emerges at minimum deviation. What is the angle of incidence upon AC ?

What is the smallest angle of incidence upon AC for which the ray can still emerge at AB ? If the angle of incidence upon AC is made zero, what will be the whole deviation of the ray? (*L.*)

9. Draw a labelled diagram of a spectrometer set up for studying the deviation of light through a triangular prism. Describe how you would adjust the instrument and use it to find the refractive index of the prism material.

Indicate briefly how you would show that the radiation from an arc lamp is not confined to the visible spectrum. (*L.*)

10. How would you investigate the way in which the deviation of a ray of light by a triangular glass prism varies with the angle of incidence on the first face of the prism? What result would you expect to obtain?

The deviation of a ray of light incident on the first face of a 60° glass prism at an angle of 45° is 40° . Find the angle which the emergent ray makes with the normal to the second face of the prism and determine, preferably by graphical construction, the refractive index of the glass of the prism. (*L.*)

11. A prism has angles of 45° , 45° , and 90° and all three faces polished. Trace the path of a ray entering one of the smaller faces in a direction parallel to the larger face and perpendicular to the prism edges. Assume 1.5 for the refractive index.

If you had two such prisms how would you determine by a simple pin or ray-box method the refractive index of a liquid available only in small quantity? (*L.*)

12. Under what circumstances does total internal reflection occur? Show that a ray of light incident in a principal section of an equilateral glass prism of refractive index 1.5, can only be transmitted after two refractions at adjacent faces if the angle of incidence on the prism exceeds a certain value. Find this limiting angle of incidence. (*W.*)

13. Draw a graph showing, in a general way, how the deviation of a ray of light when passed through a triangular prism depends on the angle of incidence.

You are required to measure the refractive index of glass in the form of a prism by means of a spectrometer provided with a vertical slit. Explain how you would level the spectrometer table and derive the formula from which you would calculate the refractive index. (You are not required to explain any other adjustment of the apparatus nor to explain how you would find the refracting angle of the prism.) (*L.*)

chapter twenty

Dispersion. Spectra

Spectrum of White Light

IN 1666, NEWTON made a great scientific discovery. He found that sunlight, or white light, was made up of different colours, consisting of red, orange, yellow, green, blue, indigo, violet. Newton made a small hole in a shutter in a darkened room, and received a white circular patch of sunlight on a screen S in the path of the light, Fig. 20.1 (i). But on interposing a glass prism between the hole and the screen he observed

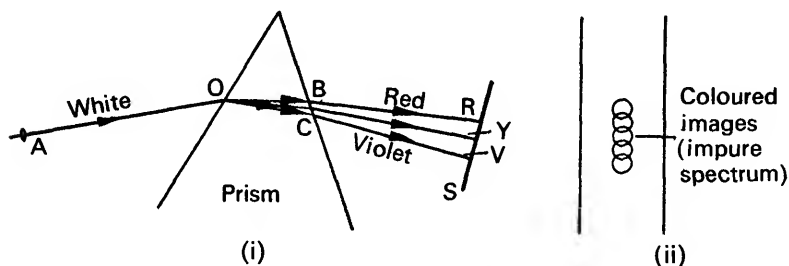


FIG. 20.1. Impure spectrum.

a series of overlapping coloured patches in place of the white patch, the total length of the coloured images being several times their width, Fig. 20.1 (ii). By separating one colour from the rest, Newton demonstrated that the colours themselves could not be changed by refraction through a prism, and he concluded that the colours were not *introduced* by the prism, but were components of the white light. The *spectrum* (colours) of white light consists of red, orange, yellow, green, blue, indigo, and violet, and the separation of the colours by the prism is known as *dispersion*.

The red rays are the least deviated by the prism, and the violet rays are the most deviated, as shown in the exaggerated sketch of Fig. 20.1 (i). Since the angle of incidence at O in the air is the same for the red and violet rays, and the angle of refraction made by the red ray OB in the glass is greater than that made by the violet ray OC, it follows from $\sin i / \sin r$ that the refractive index of the prism material for red light is less than for violet light. Similarly, the refractive index for yellow light lies between the refractive index values for red and violet light (see also p. 458).

Production of Pure Spectrum

Newton's spectrum of sunlight is an *impure spectrum* because the different coloured images overlap, Fig. 20.1 (ii). A *pure spectrum* is one in which the different coloured images contain light of one colour only, i.e., they are monochromatic images. In order to obtain a pure spectrum (i) the white light must be admitted through a very narrow opening, so as to assist in the reduction of the overlapping of the images, (ii) the beams of coloured rays emerging from the prism must be parallel, so that each beam can be brought to a separate focus.

The spectrometer can be used to provide a pure spectrum. The collimator slit is made very narrow, and the collimator C and the telescope T are both adjusted for parallel light, Fig. 20.2. A bright source

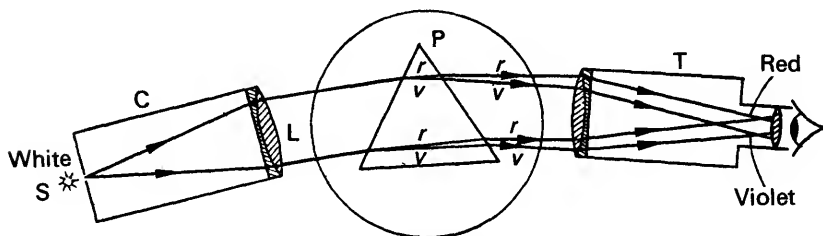


FIG. 20.2. Pure spectrum.

of white light, S, is placed near the slit, and the prism P is usually set in the minimum deviation position for yellow light, although this is not essential. The rays refracted through P are now separated into a number of different coloured parallel beams of light, each travelling in slightly different directions, and the telescope brings each coloured beam to a separate focus. A pure spectrum can now be seen through T, consisting of a series of monochromatic images of the slit.

If only one lens, L, is available, the prism P *must* be placed in the minimum deviation position for yellow light in order to obtain a fairly pure spectrum, Fig. 20.3. The prism is then also approximately in the

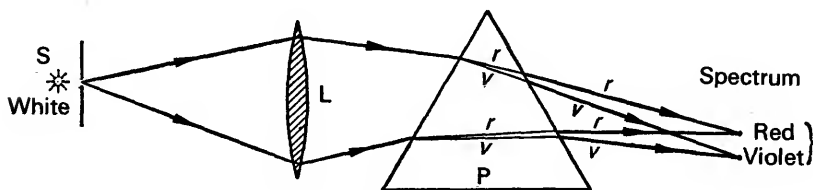


FIG. 20.3. Fairly pure spectrum.

minimum deviation position for the various colours in the incident convergent beam, and hence the rays of one colour are approximately deviated by the same amount by the prism, thus forming an image of the slit S at roughly the same place.

Infra-red and ultra-violet rays. In 1800 HERSCHEL discovered the existence

small, it follows that r_2 is a small angle. The angle of emergence i_2 is thus also small, and since $\sin i_2/\sin r_2 = n$ we may state that $i_2/r_2 = n$, or

$$i_2 = nr_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{ii})$$

The deviation, d , of the ray on passing through the prism is given by $d = (i_1 - r_1) + (i_2 - r_2)$. Substituting for i_1 and i_2 from (i) and (ii),

$$\therefore d = nr_1 - r_1 + nr_2 - r_2 = n(r_1 + r_2) - (r_1 + r_2)$$

$$\therefore d = (n - 1)(r_1 + r_2)$$

But $r_1 + r_2 = A$

$$\therefore \mathbf{d} = (n - 1) \mathbf{A} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

This is the magnitude of the deviation produced by a *small-angle* prism for *small* angles of incidence. If A is expressed in radians, then d is in radians; if A is expressed in degrees, then d is in degrees. If $A = 6^\circ$ and $n = 1.6$ for yellow light, the deviation d of that colour for small angles of incidence is given by $d = (1.6 - 1) 6^\circ = 3.6^\circ$. It will be noted that the deviation is independent of the magnitude of the small angle of incidence on the prism.

Dispersion by Small-angle Prism

We have already seen from Newton's experiment that the colours in a beam of white light are separated by a glass prism into red, orange, yellow, green, blue, indigo, violet, so that the emergent light is no longer white but coloured. The separation of the colours by the prism is known generally as the phenomenon of *dispersion*, and the *angular dispersion* between the red and blue emergent rays, for example, is defined as the *angle* between these two rays. Thus, in Fig. 20.5, θ is the angular dispersion between the red and blue rays. Of course, the angular dispersion is also equal to the *difference in deviation* of the two colours produced by the prism; and since we have already derived the expression $d = (n - 1) A$ for the deviation of monochromatic light by a small-angle prism we can obtain the angular dispersion between any two colours.

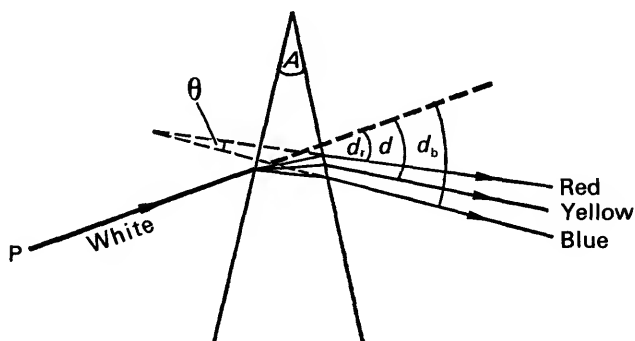


FIG. 20.5. Dispersion.

Suppose d_b, d_r are the respective deviations of the blue and red light when a ray of white light is incident at a small angle on a prism of small angle A , Fig. 20.5. Then, if n_b, n_r are the refractive indices of the prism material for blue and red light respectively,

$$d_b = (n_b - 1) A,$$

and
$$d_r = (n_r - 1) A.$$

$$\therefore \text{angular dispersion, } d_b - d_r = (n_b - 1) A - (n_r - 1) A$$

$$\therefore d_b - d_r = (n_b - n_r) A \quad (2)$$

For a particular crown glass, $n_b = 1.521, n_r = 1.510$. Thus if $A = 8^\circ$, the angular dispersion between the blue and red colours =

$$d_b - d_r = (n_b - n_r) A = (1.521 - 1.510) 8^\circ = 0.09^\circ$$

The *mean deviation* of the white light by the prism is commonly chosen as the deviation of the *yellow* light, since this is the colour approximately in the middle of the spectrum; the mean refractive index of a material is also specified as that for yellow light. Now the deviation, d , of monochromatic light is given by $d = (n - 1) A$, from equation (1), and unless otherwise stated, the magnitudes of d and n will be understood to be those for yellow light when these symbols contain no suffixes. If $n_b = 1.521$ and $n_r = 1.510$, then *approximately* the refractive index, n , for yellow light is the average of n_b and n_r , or $\frac{1}{2}(1.521 + 1.510)$; thus $n = 1.515$. Hence if the prism has an angle of 8° , the mean deviation, d , $= (n - 1) A = (1.515 - 1) 8^\circ = 4.1^\circ$.

Dispersive Power

The *dispersive power*, ω , of the material of a small-angle prism for blue and red rays may be defined as the ratio

$$\omega = \frac{\text{angular dispersion between blue and red rays}}{\text{mean deviation}} \quad (3)$$

The dispersive power depends on the material of the prism. As an illustration, suppose that a prism of angle 8° is made of glass of a type X, say, and another prism of angle 8° is made of glass of a type Y.

	n_b	n_r	n
Crown glass, X	1.521	1.510	1.515
Flint glass, Y	1.665	1.645	1.655

Further, suppose the refractive indices of the two materials for blue red, and yellow light are those shown in the above table.

For a small angle of incidence on the prism of glass X, the angular dispersion

$$\begin{aligned}
 &= d_b - d_r = (n_b - 1) A - (n_r - 1) A \\
 &= (n_b - n_r) A = (1.521 - 1.510) 8^\circ = 0.09^\circ \quad . \quad (i)
 \end{aligned}$$

$$\text{The mean deviation, } d = (n - 1) A = (1.515 - 1) 8^\circ = 4.1^\circ \quad . \quad (ii)$$

$$\therefore \text{ dispersive power, } \omega = \frac{0.09}{4.1} = 0.021 \quad . \quad . \quad . \quad (iii)$$

Similarly, for the prism of glass Y,

$$\text{angular dispersion} = (n_b - n_r) A = (1.665 - 1.645) 8 = 0.16^\circ$$

$$\text{and mean deviation} = (n - 1) A = 1.655 - 1) 8 = 5.24^\circ$$

$$\therefore \text{ dispersive power} = \frac{0.16}{5.24} = 0.03 \quad . \quad . \quad . \quad (iv)$$

From (iii) and (iv), it follows that the dispersive power of glass Y is about 1.5 times as great as that of glass X.

General Formula for Dispersive Power

We can now derive a general formula for dispersive power, ω , which is independent of angles. From equation 3, it follows that

$$\omega = \frac{d_b - d_r}{d}$$

as d_b, d_r, d denote the deviations of blue, red, and yellow light respectively.

$$\text{But } d_b - d_r = (n_b - 1) A - (n_r - 1) A = (n_b - n_r) A$$

$$\text{and } d = (n - 1) A.$$

$$\therefore \omega = \frac{d_b - d_r}{d} = \frac{(n_b - n_r) A}{(n - 1) A}$$

$$\therefore \omega = \frac{n_b - n_r}{n - 1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

From this formula, it can be seen that (i) ω depends only on the material of the prism and is independent of its angle, (ii) ω is a number and has therefore no units. In contrast to "dispersive power", it should be noted that "dispersion" is an angle, and that its magnitude depends on the angle A of the prism and the two colours concerned, for $d_b - d_r = \text{dispersion} = (n_b - n_r) A$.

Achromatic Prisms

We have seen that a prism separates the colours in white light. If a prism is required to deviate white light without dispersing it into colours, two prisms of different material must be used to eliminate the dispersion, as shown in Fig. 20.6. The prism P is made of crown glass, and causes dispersion between the red and blue in the incident white light. The prism Q is inverted with respect to P, and with a suitable choice of its angle A' (discussed fully later), the red and blue rays incident on it can be made to emerge in *parallel* directions. If the rays are viewed the eye-lens brings them to a focus at the same place on the retina, and hence

the colour effect due to red and blue rays is eliminated. The dispersion of the other colours in white light still remains, but most of the colour effect is eliminated as the red and blue rays are the "outside" (extreme) rays in the spectrum of white light.

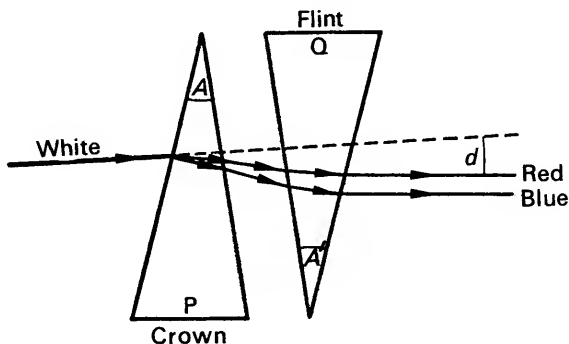


FIG. 20.6. Achromatic prisms.

Prisms which eliminate dispersion between two colours, blue and red say, are said to be *achromatic* prisms for those colours. Suppose n_b, n_r are the refractive indices of crown glass for blue and red light, and A is the angle of the crown glass prism P. Then, from p. 458,

$$\text{dispersion} = (n_b - n_r) A \quad \dots \quad (i)$$

If n'_b, n'_r are the refractive indices of flint glass for blue and red light and A' is the angle of the flint glass prism Q, then similarly,

$$\text{dispersion} = (n'_b - n'_r) A' \quad \dots \quad (ii)$$

Now prism P produces its dispersion in a "downward" direction since a prism bends rays towards its base, Fig. 20.6, and prism Q produces its dispersion in an "upward" direction. For achromatic prisms, therefore, the dispersions produced by P and Q must be equal.

$$\therefore (n_b - n_r) A = (n'_b - n'_r) A' \quad \dots \quad (5)$$

Suppose P has an angle of 6° . Then, using the refractive indices for n_b, n_r, n'_b, n'_r in the table on p. 458, it follows from (5) that the angle A' is given by

$$(1.521 - 1.510) 6^\circ = (1.665 - 1.645) A'$$

$$\text{Thus} \quad A' = \frac{0.011}{0.02} \times 6^\circ = 3.3^\circ$$

Deviation Produced by Achromatic Prisms

Although the colour effects between the red and blue rays are eliminated by the use of achromatic prisms, it should be carefully noted that the incident light beam, as a whole, has been deviated. This angle of deviation, d , is shown in Fig. 20.6, and is the angle between the incident and emergent beams. The deviation of the mean or yellow light by prism P is given by $(n - 1) A$, and is in a "downward" direction.

Since the deviation of the yellow light by the prism Q is in an opposite direction, and is given by $(n' - 1) A'$, the net deviation, d , is given by

$$d = (n - 1) A - (n' - 1) A'.$$

Using the angles 6° and 3.3° obtained above, with $n = 1.515$ and $n' = 1.655$,

$$d = (1.515 - 1) 6^\circ - (1.655 - 1) 3.3^\circ = 0.93^\circ.$$

Direct-vision Spectroscope

The direct-vision spectroscope is a simple instrument used for examining the different colours in the spectrum obtained from a glowing gas in a flame or in a discharge tube. It contains several crown and flint prisms cemented together, and contained in a straight tube having lenses which constitute an eye-piece. The tube is pointed at the source of light examined, when various colours are seen on account of the dispersion produced by the prisms, Fig. 20.7.

In practice, the direct-vision spectroscope contains several crown and flint glass prisms, but for convenience suppose we consider two such prisms, as in Fig. 20.7. For "direct vision", the *net deviation of the mean (yellow) ray* produced by the prisms must be zero. Thus the mean deviation caused by the crown glass prism in one direction must be equal to that caused by the flint glass prism in the opposite direction. Hence, with the notation already used, we must have

$$(n - 1) A = (n' - 1) A'.$$

Suppose $A = 6^\circ$, $n = 1.515$, $n' = 1.655$. Then A' is given by

$$(1.515 - 1) 6^\circ = (1.655 - 1) A'$$

$$\therefore A' = \frac{0.515}{0.655} \times 6^\circ = 4.7^\circ$$

The *net dispersion* of the blue and red rays is given by

$$\begin{aligned} & (n_b - n_r) A - (n'_b - n'_r) A' \\ &= (1.521 - 1.510) 6^\circ - (1.665 - 1.645) 4.7^\circ \\ &= 0.066 - 0.094 = -0.028^\circ. \end{aligned}$$

The minus indicates that the net dispersion is produced in a "blue-upward" direction, as the dispersion of the flint glass prism is greater than that of the crown glass prism.

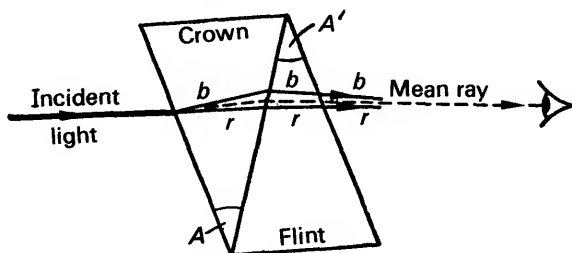


FIG. 20.7. Dispersion; but no deviation of mean ray.

SPECTRA

The Importance of the Study of Spectra

The study of the wavelengths of the radiation from a hot body comes under the general heading of *Spectra*. The number of spectra of elements and compounds which have been recorded runs easily into millions, and it is worth while stating at the outset the main reasons for the interest in the phenomenon.

It is now considered that an atom consists of a nucleus of positive electricity surrounded by electrons moving in various orbits, and that a particular electron in an orbit has a definite amount of energy. In certain circumstances the electron may jump from this orbit to another, where it has a smaller amount of energy. When this occurs radiation is emitted, and the energy in the radiation is equal to the difference in energy of the atom between its initial and final states. The displacement of an electron from one orbit to another occurs when a substance is raised to a high temperature, in which case the atoms present collide with each other very violently. Light of a definite wavelength will then be emitted, and will be characteristic of the electron energy changes in the atom. There is usually more than one wavelength in the light from a hot body (iron has more than 4,000 different wavelengths in its spectrum), and each wavelength corresponds to a change in energy between two orbits. A study of spectra should therefore reveal much important information concerning the structure and properties of atoms.

Every element has a unique spectrum. Consequently a study of the spectrum of a substance enables its composition to be readily determined. *Spectroscopy* is the name given to the exact analysis of mixtures or compounds by a study of their spectra, and the science has developed to such an extent that the presence in a substance of less than a millionth of a milligram of sodium can be detected.

Types of emission spectra. There are three different types of spectra, which are easily recognised. They are known as (a) *line spectra*, (b) *band spectra*, (c) *continuous spectra*.

(a) *Line spectra.* When the light emitted by the atoms of a glowing substance (such as vaporised sodium or helium gas) is examined by a prism and spectrometer, lines of various wavelengths are obtained. These lines, it should be noted, are images of the narrow slit of the spectrometer on which the light is incident. The spectra of hydrogen, Fig. 20.8, and helium are line spectra, and it is generally true that line spectra are obtained from *atoms*.

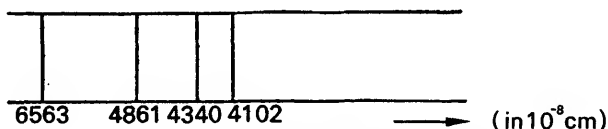


FIG. 20.8. Visible line spectra of hydrogen.

(b) *Band spectra.* Band spectra are obtained from *molecules*, and consist of a series of bands each sharp at one end but "fading" at the other end, Fig.

20.9. The term "fluting" is often used to describe the way in which the bands are spaced. Careful examination reveals that the bands are made up of numerous fine lines very close to each other. Two examples of band spectra are those usually obtained from nitrogen and oxygen.

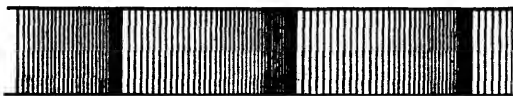


FIG. 20.9. Diagrammatic representation of band spectra.

(c) *Continuous spectra.* The spectrum of the sun is an example of a continuous spectrum, and, in general, the latter are obtained from *solids and liquids*. In these states of matter the atoms and molecules are close together, and electron orbital changes in a particular atom are influenced by neighbouring atoms to such an extent that radiations of all different wavelengths are emitted. In a gas the atoms are comparatively far apart, and each atom is uninfluenced by any other. The gas therefore emits radiations of wavelengths which result from orbital changes in the atom due solely to the high temperature of the gas, and a line spectrum is obtained. When the temperature of a gas is decreased and pressure applied so that the liquid state is approached, the line spectrum of the gas is observed to broaden out considerably.

Production of spectra. In order to produce its spectrum the substance under examination must be heated to a high temperature. There are four main methods of *excitation*, as the process is called, and spectra are classified under the method of their production.

(a) *Flame spectra.* The temperature of a Bunsen flame is high enough to vaporise certain solids. Thus if a piece of platinum wire is dipped into a sodium salt and then placed in the flame, a vivid yellow colour is obtained which is characteristic of the element sodium. This method of excitation can only be used for a limited number of metals, the main class being the alkali and alkaline earth metals such as sodium, potassium, lithium, calcium, and barium. The line spectra produced in each case consist of lines of different colours, but some lines have a greater intensity than others. Thus sodium is characterised by two prominent yellow lines barely distinguishable in a small spectroscope, and lithium by a prominent green line.

(b) *Spark spectra.* If metal electrodes are connected to the secondary of an induction coil and placed a few millimetres apart, a spark can be obtained which bridges the gap. It was discovered that a much more intense and violent spark could be obtained by placing a capacitor in parallel with the gap. This spark is known as a *condensed spark*. The solid under investigation forms one of the electrodes, and is vaporised at the high temperature obtained.

(c) *Arc spectra.* This is the method most used in industry. If two metal rods connected to a d.c. voltage supply are placed in contact with each other and then drawn a few millimetres apart, a continuous spark, known as an arc, is obtained across the gap. The arc is a source of very high temperature, and therefore vaporises substances very readily. In practice the two rods are placed in a vertical position, and a small amount of the substance investigated is placed on the lower rod.

(d) *Discharge-tube spectra.* If a gas is contained at low pressure inside a tube having two aluminium electrodes and a high a.c. or d.c. voltage is applied to the gas, a "discharge" occurs between the electrodes and the gas

becomes luminous. This is the most convenient method of examining the spectra of gases. The luminous neon gas in a discharge tube has a reddish colour, while mercury vapour is greenish-blue.

Absorption Spectra. Kirchhoff's Law

The spectra just discussed are classified as *emission spectra*. There is another class of spectra known as *absorption spectra*, which we shall now briefly consider.

If light from a source having a continuous spectrum is examined after it has passed through a sodium flame, the spectrum is found to be crossed by a dark line; this dark line is in the position corresponding to the bright line emission spectrum obtained with the sodium flame alone. The continuous spectrum with the dark line is naturally characteristic of the absorbing substance, in this case sodium, and it is known as an *absorption spectrum*. An absorption spectrum is obtained when red glass is placed in front of sunlight, as it allows only a narrow band of red rays to be transmitted.

KIRCHHOFF's investigations on absorption spectra in 1855 led him to formulate a simple law concerning the emission and absorption of light by a substance. This states: *A substance which emits light of a certain wavelength at a given temperature can also absorb light of the same wavelength at that temperature*. In other words, a good emitter of a certain wavelength is also a good absorber of that wavelength. From Kirchhoff's law it follows that if the radiation from a hot source emitting a continuous spectrum is passed through a vapour, the absorption spectrum obtained is deficient in those wavelengths which the vapour would emit if it were raised to the same high temperature. Thus if a sodium flame is observed through a spectrometer in a darkened room, a bright yellow line is seen; if a strong white arc light, richer in yellow light than the sodium flame, is placed behind the flame, a dark line is observed in the place of the yellow line. The sodium absorbs yellow light from the white light, and re-radiates it in all directions. Consequently there is less yellow light in front of the sodium flame than if it were removed, and a dark line is thus observed.

Fraunhofer Lines

In 1814 FRAUNHOFER noticed that the sun's spectrum was crossed by many hundreds of dark lines. These *Fraunhofer lines*, as they are called, were mapped out by him on a chart of wavelengths, and the more prominent were labelled by the letters of the alphabet. Thus the dark line in the blue part of the spectrum was known as the *F* line, the dark line in the yellow part as the *D* line, and the dark line in the red part as the *C* line.

The Fraunhofer lines indicate the presence in the sun's atmosphere of certain elements in a vaporised form. The vapours are cooler than the central hot portion of the sun, and they absorb their own characteristic wavelengths from the sun's continuous spectrum. Now every element

has a characteristic spectrum of wavelengths. Accordingly, it became possible to identify the elements round the sun from a study of the wavelengths of the Fraunhofer (dark) lines in the sun's spectrum, and it was then found that hydrogen and helium were present. This was how helium was first discovered. The *D* line is the yellow sodium line.

The incandescent gases round the sun can be seen as flames many miles high during a total eclipse of the sun, when the central portion of the sun is cut off from the observer. If the spectrum of the sun is observed just before an eclipse takes place, a continuous spectrum with Fraunhofer lines is obtained, as already stated. At the instant when the eclipse becomes total, however, bright emission lines are seen in exactly the same position as those previously occupied by the Fraunhofer lines, and they correspond to the emission spectra of the vapours alone. This is an illustration of Kirchhoff's law, p. 464.

Measurement of Wavelengths by Spectrometer

As we shall discuss later (p. 690) the light waves produced by different colours are characterised by different *wavelengths*. Besides measuring refractive index, the spectrometer can be adapted for measuring unknown wavelengths, corresponding to the lines in the spectrum of a glowing gas in a discharge tube, for example.

A prism is first placed on the spectrometer table in the minimum deviation position for yellow (sodium) light, thus providing a reference position for the prism in relation to incident light from the collimator. The source of yellow light is now replaced by a helium discharge tube, which contains helium at a very low pressure, glowing as a result of the high voltage placed across the tube. Several bright lines of various colours can now be observed through the telescope (they are differently coloured images of the slit), and the *deviation*, θ , of each of the lines is obtained by rotating the telescope until the image is on the cross-wires,

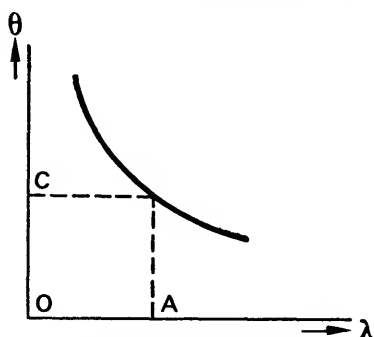


FIG. 20.10. Measurement of wavelength by spectrometer.

and then noting the corresponding reading on the circular graduated scale. Since the wavelengths, λ , of the various lines in the helium spectrum are known very accurately from tables, a graph can now be plotted between θ and λ . The helium discharge tube can then be replaced by a hydrogen or mercury discharge tube, and the deviations due to other lines of known wavelength obtained. In this way a *calibration curve* for the spectrometer can be obtained, Fig. 20.10.

The wavelength due to a line *Q* in the spectrum of an unknown glowing gas can now be easily derived. With the prism still in the minimum deviation position for yellow light, the deviation, θ , of *Q* is

measured. If this angle corresponds to C in Fig. 20.10, the wavelength λ is OA.

EXAMPLES

1. Show that when a ray of light passes nearly normally through a prism of small angle α and refractive index n , the deviation δ is given by $\delta = (n - 1)\alpha$. A parallel beam of light falls normally upon the first face of a prism of small angle. The portion of the beam which is refracted at the second surface is deviated through an angle of $1^\circ 35'$, and the portion which is reflected at the second surface and emerges again at the first surface makes an angle of $8^\circ 9'$ with the incident beam. Calculate the angle of the prism and the refractive index of the glass. (C.)

First part. See text.

Second part. Let θ = angle of prism, n = the refractive index, and RH the ray incident normally on the face AN, striking the second face at K, Fig. 20.11.

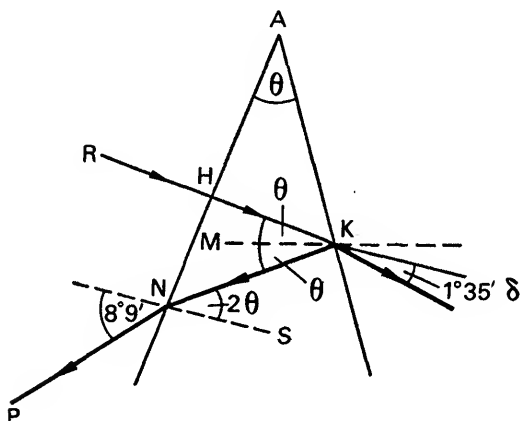


FIG. 20.11. Example.

Then the angle of incidence at $K = \theta$, and angle $HKN = 2\theta$. By drawing the normal NS at N , which is parallel to HK , it can be seen that angle $KNS = 2\theta$. The angle of emergence from the prism $= 8^\circ 9'$ since the incident beam was normal to AN .

The angle of deviation, δ , of the beam by the prism is given by

$$\therefore \delta = (n - 1)\theta \quad \therefore 1^\circ 35' = (n - 1)\theta \quad \dots \quad (i)$$

$$\text{For refraction at } N, n = \frac{\sin 8^\circ 9'}{\sin 2\theta}$$

Since the angles concerned are small,

$$n = \frac{8^\circ 9'}{2\theta} \quad \dots \quad (ii)$$

where θ is in degrees.

From (ii), $\theta = \frac{8^\circ 9'}{2n}$; substituting in (i),

$$\therefore 1^\circ 35' = (n - 1) \frac{8^\circ 9'}{2n}$$

$$\therefore \frac{n - 1}{2n} = \frac{95}{489}$$

$$\therefore 489n - 489 = 190n$$

$$\therefore n = 1.63$$

$$\therefore \theta = \frac{8^\circ 29'}{2n} = \frac{8^\circ 29'}{3.26} = 2^\circ 30'$$

2. Define dispersive power. The following table gives the refractive indices of crown and flint glass for three lines of the spectrum.

	C	D	F
Crown	1.514	1.517	1.523
Flint	1.644	1.650	1.664

Calculate the refracting angle of a flint glass prism which, when combined with a crown glass prism of refracting angle 5° , produces a combination that does not deviate the light corresponding to the D line. What separation of the rays corresponding to the C and F lines will such a compound prism produce? (L.)

For definition, see text.

The D line corresponds to the mean, or yellow, ray, the F and C lines to the blue and red rays respectively. Let n' , n = the refractive indices for crown and flint glass respectively, A' , A = the corresponding angles of the prisms.

For no deviation $(n'_D - 1)A' - (n_D - 1)A = 0$,

$$\therefore (1.517 - 1)5^\circ - (1.650 - 1)A = 0$$

$$\therefore A = \frac{0.517}{0.650} \times 5 = 3.99^\circ$$

The separation of the F and C lines

$$= (n_F - n_C)A - (n'_F - n'_C)A'$$

$$= (1.664 - 1.644)3.99^\circ - (1.523 - 1.514)5^\circ$$

$$= 0.0798^\circ - 0.045^\circ = 0.0348^\circ$$

3. Prove that for a prism of small angle A the deviation of a ray of light is $(n - 1)A$, provided that the angle of incidence also is small. A crown glass prism of refracting angle 6° is to be achromatised for red and blue light with a flint glass prism. Using the data below and the formula above find (a) the angle of the flint glass prism, (b) the mean deviation.

	Crown glass	Flint glass
n red	1.513	1.645
n blue	1.523	1.665

(N.)

First part. See text.

Second part. Let A = the angle of the flint prism, n' , n = the refractive indices of the crown and flint glass respectively. For achromatism,

$$(n'_b - n'_r) 6^\circ = (n_b - n_r) A$$

$$\therefore (1.523 - 1.513) 6^\circ = 1.665 - 1.645) A$$

$$\therefore A = \frac{0.010}{0.020} \times 6^\circ = 3^\circ$$

$$\text{The mean refractive index, } n', \text{ for crown glass} = \frac{1.523 + 1.513}{2} = 1.518$$

$$\text{and mean refractive index } n, \text{ for flint glass} = \frac{1.665 + 1.645}{2} = 1.655$$

$$\therefore \text{deviation of mean ray} = (n' - 1) 6^\circ - (n - 1) 3^\circ$$

$$= (1.518 - 1) 6^\circ - (1.655 - 1) 3^\circ = 1.043^\circ$$

EXERCISES 20

1. Write down the formula for the deviation of a ray of light through a prism of small angle A which has a refractive index n for the colour concerned. Using the following table, calculate the deviation of (i) red light, (ii) blue light, (iii) yellow light through a flint glass prism of refracting angle 4° , and through a crown glass prism of refracting angle 6° .

	Crown glass	Flint glass
n red	1.512	1.646
n blue	1.524	1.666

2. Using the above data, calculate the *dispersive powers* of crown glass and flint glass.

3. Explain how it is possible with two prisms to produce dispersion without mean deviation. A prism of crown glass with refracting angle of 5° and mean refractive index 1.51 is combined with one flint glass of refractive index 1.65 to produce no mean deviation. Find the angle of the flint glass prism. The difference in the refractive indices of the red and blue rays in crown glass is 0.0085 and in flint glass 0.0162. Find the inclination between the red and blue rays which emerge from the composite prism. (L .)

4. Draw a ray diagram showing the passage of light of two different wavelengths through a prism spectrometer. Why is it that such a spectrometer is almost invariably used with (a) a very narrow entrance slit, (b) parallel light passing through the prism, (c) the prism set at, or near, minimum deviation?

A spectrometer is used with a *small angle* prism made from glass which has a refractive index of 1.649 for the blue mercury line and 1.631 for the green mercury line. The collimator lens and the objective of the spectrometer both have a focal length of 30 cm. If the angle of the prism is 0.1 radian what is the spacing of the centres of the blue and green mercury lines in the focal plane of the objective, and what maximum slit width may be used without the lines overlapping? The effect of diffraction need not be considered. (O . & C .)

5. A glass prism of refracting angle 6.0° and of material of refractive index 1.50 is held with its refracting angle downwards alongside another prism of angle 4.0° which has its refracting angle pointing upwards. A narrow parallel beam of yellow light is incident nearly normally on the first prism, passes through both prisms, and is observed to emerge parallel to its original direction. Calculate the refractive index of the material of the second prism. If white light were used and the glasses of the two prisms were very different in their power to disperse light, describe very briefly what would be seen on a white screen placed at right angles to the emergent light. (C.)

6. A ray of monochromatic light is incident at an angle i on one face of a prism of refracting angle A of glass of refractive index n and is transmitted. The deviation of the ray is D .

Considering only rays incident on the side of the normal away from the refracting angle, sketch graphs on the same set of axes showing how D varies with i when (a) A is about 60° , (b) A is very small.

From first principles derive an expression for D when i and A are both very small angles. (N.)

7. Distinguish between emission spectra and absorption spectra. Describe the spectrum of the light emitted by (i) the sun, (ii) a car headlamp fitted with yellow glass, (iii) a sodium vapour street lamp.

What are the approximate wavelength limits of the visible spectrum? How would you demonstrate the existence of radiations whose wavelengths lie just outside these limits? (O. & C.)

8. State what is meant by *dispersion* and describe, with diagrams, the principle of (i) an achromatic and (ii) a direct-vision prism.

Derive an expression for the refractive index of the glass of a *narrow* angle prism in terms of the angle of minimum deviation and the angle of the prism. If the refractive index of the glass of refracting angle 8° is 1.532 and 1.514 for blue and red light respectively, determine the angular dispersion produced by the prism. (L.)

9. Describe the processes which lead to the formation of numerous dark lines (Fraunhofer lines) in the solar spectrum. Explain why the positions of these lines in the spectrum differ very slightly when the light is received from opposite ends of an equatorial diameter of the sun. (N.)

10. Describe with the aid of diagrams what is meant by *dispersion* and *deviation* by a glass prism. Derive a formula for the deviation D produced by a glass prism of small refracting angle A for small angles of incidence. Sketch the graph showing how the deviation varies with angle of incidence for a beam of light striking such a prism, and on the same axes indicate what would happen with a prism of much larger refracting angle but of material of the same index of refraction. (C.)

11. Describe the optical system of a simple prism spectrometer. Illustrate your answer with a diagram showing the paths through the spectrometer of the pencils of rays which form the red and blue ends of the spectrum of a source of white light. (Assume in your diagram that the lenses are achromatic.)

The prism of a spectrometer has a refracting angle of 60° and is made of glass whose refractive indices for red and violet are respectively 1.514 and 1.530. A white source is used and the instrument is set to give minimum deviation for red light. Determine (a) the angle of incidence of the light on the prism, (b) the angle of emergence of the violet light, (c) the angular width of the spectrum. (N.)

12. Calculate the angle of a crown glass prism which makes an achromatic combination for red and blue light with a flint glass prism of refracting angle 4° . What is the mean deviation of the light by this combination? Use the data given in question 1.

13. Describe and give a diagram of the optical system of a spectrometer. What procedure would you adopt when using the instrument to measure the refractive index of the glass of a prism for sodium light? What additional observations would be necessary in order to determine the dispersive power of the glass?

The refractive index of the glass of a prism for red light is 1.514 and for blue light 1.523. Calculate the difference in the velocities of the red and blue light in the prism if the velocity of light *in vacuo* is 3×10^5 kilometres per second. (N.)

14. Explain, with diagrams, how a 'pure' spectrum is produced by means of a spectrometer. What source of light may be used and what readings must be taken in order to find the dispersive power of the material of which the prism is made? (L.)

15. (a) Explain, giving a carefully drawn, labelled diagram, the function of the various parts of a spectrometer. How is it adjusted for normal laboratory use? (b) Distinguish between a continuous spectrum, an absorption spectrum, a band spectrum, and a line spectrum. State briefly how you would obtain each type with a spectrometer. (W.)

16. Describe a prism spectrometer and the adjustment of it necessary for the precise observation of the spectrum of light by a gaseous source.

Compare and contrast briefly the spectrum of sunlight and of light emitted by hydrogen at low pressure contained in a tube through which an electric discharge is passing. (L.)

chapter twenty-one

Refraction through lenses

A *lens* is a piece of glass bounded by one or two spherical surfaces. When a lens is thicker in the middle than at the edges it is called a *convex* or *converging lens*, Fig. 21.1 (i); when it is thinner in the middle than at the edges it is known as a *concave* or *diverging lens*, Fig. 21.1 (ii). Fig. 21.9, on p. 478, illustrates other types of converging and diverging lenses.

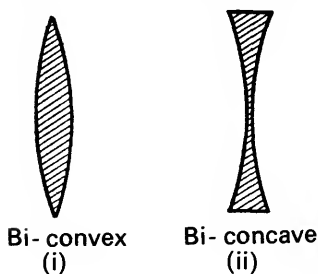


FIG. 21.1.
Converging and diverging lenses.

Lenses were no doubt made soon after the art of glass-making was discovered; and as the sun's rays could be concentrated by these curved pieces of glass they were called "burning glasses". ARISTOPHANES, in 424 B.C., mentions a burning glass. To-day, lenses are used in spectacles, cameras, microscopes, and telescopes, as well as in many other optical instruments, and they afford yet another example of the many ways in which Science is used to benefit our everyday lives.

Since a lens has a curved spherical surface, a thorough study of a lens should be preceded by a discussion of the refraction of light through a curved surface. We shall therefore proceed to consider what happens in this case, and defer a discussion of lenses until later, p. 478.

REFRACTION AT CURVED SPHERICAL SURFACE

Relation Between Object and Image Distances

Consider a curved spherical surface NP, bounding media of refractive indices n_1 , n_2 respectively, Fig. 21.2. The medium of refractive index n_1 might be air, for example, and the other of refractive index n_2 might be glass. The centre, C, of the sphere of which NP is part is the centre of curvature of the surface, and hence CP is the radius of curvature, r . The line joining C to the mid-point P of the surface is known as its *principal axis*. P is known as the *pole*.

Suppose a point object O is situated on the axis PC in the medium of refractive index n_1 . The image of O by refraction at the curved surface can be obtained by taking two rays from O. A ray OP passes straight through along PC into the medium of refractive index n_2 , since OP is normal to the surface, while a ray ON *very close* to the axis is refracted

at N along NI towards the normal CN, if we assume n_2 is greater than n_1 . Thus at the point of intersection, I, of OP and NI is the image O, and we have here the case of a real image.

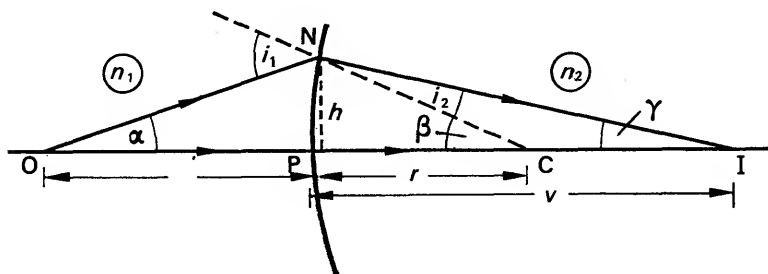


FIG. 21.2. Refraction at curved surface.

Suppose i_1, i_2 are the angles made by ON, NI respectively with the normal, CN, at N, Fig. 21.2. Then, applying " $n \sin i$ " is a constant (p. 423),

$$n_1 \sin i_1 = n_2 \sin i_2 \quad \text{. (i)}$$

But if we deal with rays from O very close to the axis OP, i_1 is small; and hence $\sin i_1 = i_1$ in radians. Similarly, $\sin i_2 = i_2$ in radians. From (i), it follows that

$$n_1 i_1 = n_2 i_2 \quad \text{. (ii)}$$

If α, β, γ are the angles with the axis made by ON, CN, IN respectively, we have

$$i_1 = \alpha + \beta, \text{ from the geometry of triangle ONC,}$$

and

$$i_2 = \beta - \gamma, \text{ from the geometry of triangle CN I.}$$

Substituting for i_1, i_2 in (ii), we have

$$\begin{aligned} n_1 (\alpha + \beta) &= n_2 (\beta - \gamma) \\ \therefore n_1 \alpha + n_2 \gamma &= (n_2 - n_1) \beta \quad \text{. (iii)} \end{aligned}$$

If h is the height of N above the axis, and N is so close to P that NP is perpendicular to OP,

$$\alpha = \frac{h}{OP}, \gamma = \frac{h}{PI}, \beta = \frac{h}{PC},$$

From (iii), using our sign convention on p. 407, we have

$$h \left(\frac{n_1}{+OP} + \frac{n_2}{+PI} \right) = h \frac{(n_2 - n_1)}{PC},$$

since O is a real object and I is a real image.

$$\therefore \frac{n_1}{OP} + \frac{n_2}{PI} = \frac{n_2 - n_1}{PC}.$$

If the object distance, OP, from P = u , the image distance, IP, from P = v , and PC = r , then

$$\frac{n_1}{u} + \frac{n_2}{v} = \frac{n_2 - n_1}{r} \quad \text{. (1)}$$

Sign Convention for Radius of Curvature

Equation (1) is the general relation between the object and image distances, u , v , from the middle or pole of the refracting surface, its radius of curvature r , and the refractive indices of the media, n_2 , n_1 . The quantity $(n_2 - n_1)/r$ is known as the **power** of the surface. If a ray is made to converge by a surface, as in Fig. 21.1, the power will be assumed *positive* in sign; if a ray is made to diverge by a surface, the power will be assumed negative. Since refractive index is a ratio of velocities (p. 421), n_1 and n_2 have no sign. $(n_2 - n_1)$ on the right side of equation (1) will be taken always as a *positive* quantity, and thus denotes the smaller refractive index subtracted from the greater refractive index. The sign convention for the radius of curvature, r , of a spherical surface is now as follows: if the surface is *convex to the less dense medium*, its radius is *positive*; if it is concave to the less dense medium, its radius is negative. We have thus to view the surface from a point in the less dense medium. In Fig. 21.3 (i), the surface A is convex to the less dense

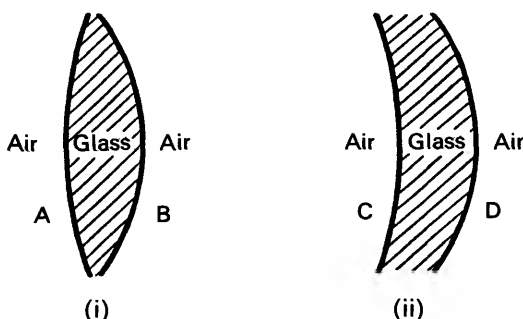


FIG. 21.3. Sign convention for radius of curvature.

medium air, and hence its radius is positive. The surface C is concave to the less dense medium air, and its radius is thus negative, Fig. 21.3 (ii). The radii of the surfaces B and D are both positive.

Special Cases

The general formula $\frac{n_1}{u} + \frac{n_2}{v} = \frac{n_2 - n_1}{r}$ can easily be remembered on account of its symmetry. The object distance u corresponds to the refractive index n_1 of the medium in which the object is situated; while the image distance v corresponds to the medium of refractive index n_2 in which the image is situated.

Suppose an object O in air is x cm from a curved spherical surface, and the image I is real and in glass of refractive index n , at a distance of y cm from the surface, Fig. 21.4 (i). Then $u = +x$, $v = +y$, $n_1 = 1$, $n_2 = n$. If the surface is convex to the less dense medium, as shown in Fig. 21.4 (i), the radius of curvature, a cm, is given by $r = +a$.

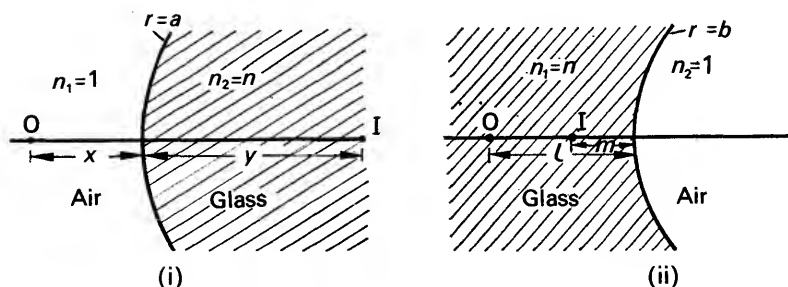


FIG. 21.4. Special cases of refraction formula.

Substituting in
$$\frac{n_1}{u} + \frac{n_2}{v} = \frac{n_2 - n_1}{r}$$

$$\therefore \frac{1}{x} + \frac{n}{y} = \frac{n - 1}{a}$$

Fig. 21.4 (ii) illustrates the case of an object O in glass of refractive index n , the surface being concave to the less dense surface air. The radius, b cm, is then given by $r = -b$. If the image I is virtual, its distance $v = -m$. If l is the distance of O, then $u = +l$.

Substituting in
$$\frac{n_1}{u} + \frac{n_2}{v} = \frac{n_2 - n_1}{r}$$

$$\therefore \frac{n}{l} + \frac{1}{-m} = \frac{n - 1}{-b}$$

If a surface is *plane*, its radius of curvature, r , is infinitely large.

Hence $\frac{n_2 - n_1}{r}$ is zero, whatever different values n_1 and n_2 may have.

Deviation of Light by Sphere

Suppose a ray AO in air is incident on a sphere of glass or a drop of water, Fig. 21.5 (i). The light is refracted at O, then reflected inside B,

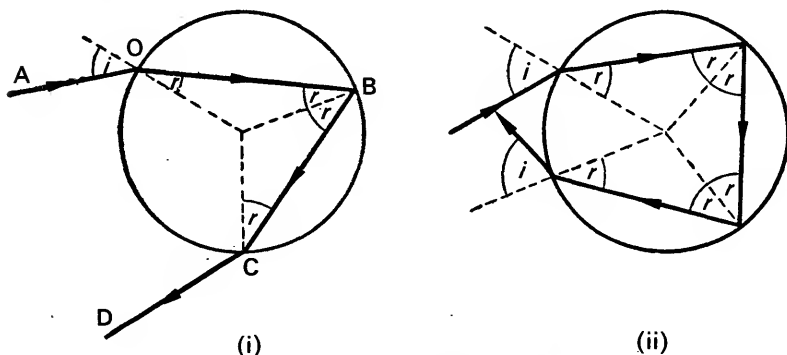


FIG. 21.5. Deviation of light by sphere.

and finally emerges into the air along CD. If i, r are the angles of incidence and refraction at O, the deviation of the light at O and C is $(i - r)$ each time; it is $(180^\circ - 2r)$ at B. The total deviation, δ , in a clockwise direction is thus given by

$$\delta = 2(i - r) + 180^\circ - 2r = 180^\circ + 2i - 4r \quad (i) \quad . \quad . \quad .$$

It can be seen that the deviation at each reflection inside the sphere is $(180^\circ - 2r)$ and that the deviation at each refraction is $(i - r)$. Thus if a ray undergoes two reflections inside the sphere, and two refractions, as shown in Fig. 21.5 (ii), the total deviation in a clockwise direction = $2(i - r) + 2(180^\circ - 2r) = 360^\circ + 2i - 6r$. After m internal reflections,

$$\text{the total deviation} = 2(i - r) + m(180^\circ - 2r).$$

The Rainbow

The explanation of the colours of the rainbow was first given by Newton about 1667. He had already shown that sunlight consisted of a mixture of colours ranging from red to violet, and that glass could disperse or separate the colours (p. 454). In the same way, he argued, water droplets in the air dispersed the various colours in different directions, so that the colours of the spectrum were seen.

The curved appearance of the rainbow was first correctly explained about 1611. It was attributed to refraction of light at a water drop, followed by reflection inside the drop, the ray finally emerging into the air as shown in Fig. 21.6. The *primary bow* is the rainbow usually seen, and is obtained by two refractions and one reflection at the drops, as in Fig. 21.6. Sometimes a *secondary bow* is seen higher in the sky, and it is formed by rays undergoing two refractions and two reflections at the drop, as in Fig. 21.6.

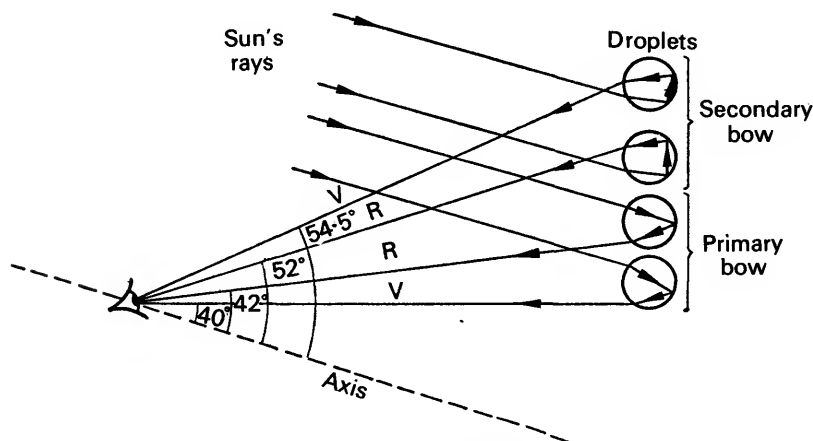


FIG. 21.6. The Rainbow.

The total deviation δ of the light when one reflection occurs in the drop, Fig. 21.6, is given by $\delta = 180^\circ + 2i - 4r$, as proved before. Now the light

emerging from the drop will be intense at those angles of incidence corresponding to the *minimum* deviation position, since a considerable number of rays have about the same deviation at the minimum value and thus emerge

almost parallel. Now for a minimum value, $\frac{d\delta}{di} = 0$.

Differentiating the expression for δ , we have $2 - 4 \frac{dr}{di} = 0$.

$$\therefore \frac{dr}{di} = \frac{1}{2}$$

But

$$\sin i = n \sin r,$$

where n is the refractive index of water.

$$\therefore \cos i = n \cos r \frac{dr}{di} = \frac{n}{2} \cos r$$

$$\therefore 4 \cos^2 i = n^2 \cos^2 r = n^2 - n^2 \sin^2 r = n^2 - \sin^2 i$$

$$= n^2 - (1 - \cos^2 i)$$

$$\therefore 3 \cos^2 i = n^2 - 1$$

$$\therefore \cos i = \sqrt{\frac{n^2 - 1}{3}} \quad \dots \dots \dots (ii)$$

The refractive index of water for red light is 1.331. Substituting this value in (ii) i can be found, and thus r is obtained. The deviation δ can then be calculated, and the acute angle between the incident and emergent red rays, which is the supplement of δ , is about 42.1° . By substituting the refractive index of water for violet light in (ii), the acute angle between the incident and emergent violet rays is found to be about 40.2° . Thus if a shower of drops is illuminated by the sun's rays, an observer standing with his back to the sun sees a brilliant red light at an angle of 42.1° with the line joining the sun to him, and a brilliant violet light at an angle of 40.2° with this line, Fig. 21.6. Since the phenomenon is the same in all planes passing through the line, the brightly coloured drops form an arc of a circle whose centre is on the line.

The secondary bow is formed by two internal reflections in the water drops, as illustrated in Fig. 21.5 (ii) and Fig. 21.6. The minimum deviation occurs when $\cos i = \sqrt{(n^2 - 1)/8}$ in this case. The acute angle between the incident and emergent red rays is then found to be about 51.8° , and that for the violet rays is found to be about 54.5° . Thus the secondary bow has red on the inside and violet on the outside, whereas the primary bow colours are the reverse, Fig. 21.6.

EXAMPLES

1. Obtain a formula connecting the distances of object and image from a spherical refracting surface. A small piece of paper is stuck on a glass sphere of 5 cm radius and viewed through the glass from a position directly opposite. Find the position of the image. Find also the position of the image formed, by the sphere, of an object at infinity. (*O.* & *C.*)

First part. See text.

Second part. Suppose *O* is the piece of paper, Fig. 21.7 (i). The refracting surface of the glass is at *P*, and $u = +10$. Now

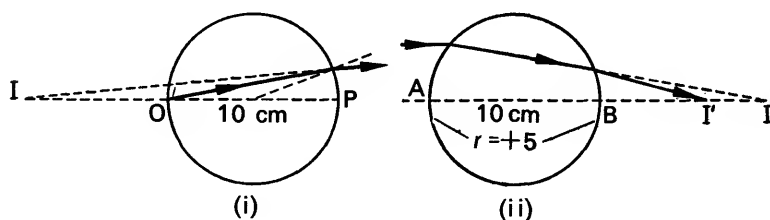


FIG. 21.7. Example

$$\frac{n_2}{v} + \frac{n_1}{u} = \frac{n_2 - n_1}{r}$$

where $n_1 = 1.5$, $n_2 = 1$, $r = +5$ p. 472 and v is the image distance from P.

Substituting,
$$\frac{1}{v} + \frac{1.5}{10} = \frac{1.5 - 1}{5}$$

$$\therefore \frac{1}{v} = 0.1 - 0.15 = -0.05$$

$$\therefore v = -20 \text{ cm.}$$

Thus the image is virtual, i.e., it is 20 cm from P on the same side as O.

Third part. Suppose I is the position of the image by refraction at the first surface, A, Fig. 21.7 (ii). Now $\frac{n_2}{v} + \frac{n_1}{u} = \frac{n_2 - n_1}{r}$, where $u = \infty$, $n_1 = 1$, $n_2 = 1.5$, $r = +5$.

$$\therefore \frac{1.5}{v} = \frac{1.5 - 1}{5}$$

$$\therefore v = 15 \text{ cm} = \text{AI, or BI} = 5 \text{ cm.}$$

I is a virtual object for refraction at the curved surface B. Since $u = -\text{BI} = -5$ cm, $n_1 = 1.5$, $n_2 = 1$, $r = +5$, it follows from

$$\frac{n_2}{v} + \frac{n_1}{u} = \frac{n_2 - n_1}{r}$$

that
$$\frac{1}{v} + \frac{1.5}{(-5)} = \frac{1.5 - 1}{5}$$

from which
$$v = 2.5 \text{ cm} = \text{BI}'.$$

2. An object is placed in front of a spherical refracting surface. Derive an expression connecting the distances from the refracting surface of the object and the image produced. The apparent thickness of a thick plano-convex lens is measured with (a) the plane face uppermost (b) the convex face uppermost, the values being 2 cm and $2\frac{2}{3}$ cm respectively. If its real thickness is 3 cm, calculate the refractive index of the glass and the radius of curvature of the convex face. (L.)

First part. See text.

Second part. With the plane face uppermost, the image I of the lowest point O is obtained by considering refraction at the plane surface D, Fig. 21.8 (i). Now

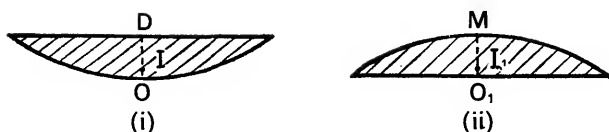


FIG. 21.8. Example

$$n = \frac{\text{real depth}}{\text{apparent depth}}$$

$$\therefore n = \frac{3}{2} = 1.5$$

With the curved surface uppermost, the image I_1 of the lowest point O_1 is obtained by considering refraction at the curved surface M, Fig. 21.8 (ii). In this case $MI_1 = v = \text{apparent thickness} = -2\frac{2}{3}$ cm, the image I_1 being virtual. Now $u = MO_1 = 3$ cm, $n_2 = 1$, $n_1 = n = 1.5$. Substituting in

$$\frac{n_2}{v} + \frac{n_1}{u} = \frac{n_2 - n_1}{r}$$

$$\text{we have} \quad \frac{1}{-2\frac{2}{3}} + \frac{1.5}{3} = \frac{1.5 - 1}{r}.$$

Simplifying,

$$r = 10 \text{ cm}.$$

REFRACTION THROUGH THIN LENSES

Converging and Diverging Lenses

At the beginning of the chapter we defined a lens as an object, usually of glass, bounded by one or two spherical surfaces. Besides the converging (convex) lens shown in Fig. 21.1 (i) on p. 471, Fig. 21.9 (i) illustrates two other types of converging lenses, which are thicker in the middle than at the edges. Fig. 84 (ii) illustrates two types of diverging (concave) lenses, a diverging lens being also shown in Fig. 21.1 (ii) on p. 471.

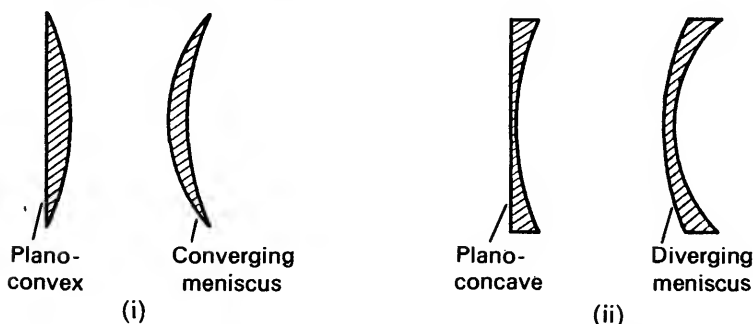


FIG. 21.9. (i). Convex (converging) lenses. (ii). Concave (diverging) lenses.

The *principal axis* of a lens is the line joining the centres of curvature of the two surfaces, and passes through the middle of the lens. Experi-

ments with a ray-box show that a thin convex lens brings an incident parallel beam of rays to a *principal focus*, F , on the other side of the lens when the beam is narrow and incident close to the principal axis, Fig. 21.10 (i). On account of the convergent beam contained with it, the convex lens is better described as a “converging” lens. If a similar parallel beam is incident on the other (right) side of the lens, it converges to a focus F' , which is at the same distance from the lens as F when the lens is thin. To distinguish F from F' the latter is called the “first principal focus”; F is known as the “second principal focus”.

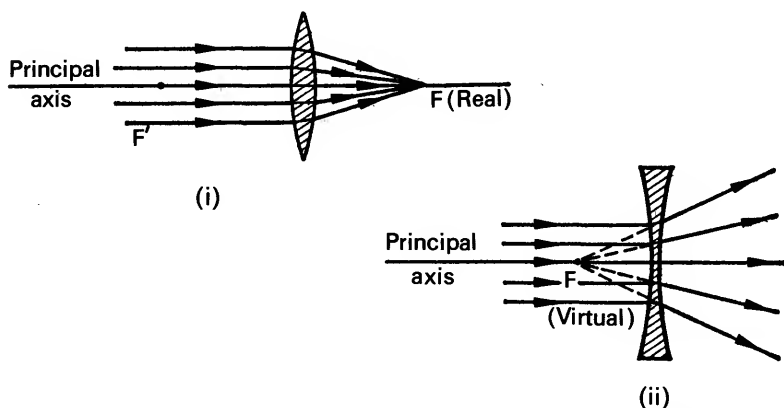


FIG. 21.10. Focus of converging (convex) and diverging (concave) lenses.

When a narrow parallel beam, close to the principal axis, is incident on a thin concave lens, experiment shows that a beam is obtained which appears to diverge from a point F on the same side as the incident beam, Fig. 21.10 (ii). F is known as the principal “focus” of the concave lens. Since a divergent beam is obtained, the concave lens is better described as a “diverging” lens.

Explanation of Effects of Lenses

A thin lens may be regarded as made up of a very large number of *small-angle prisms* placed together, as shown in the exaggerated sketches of Fig. 21.11. If the spherical surfaces of the various truncated prisms are

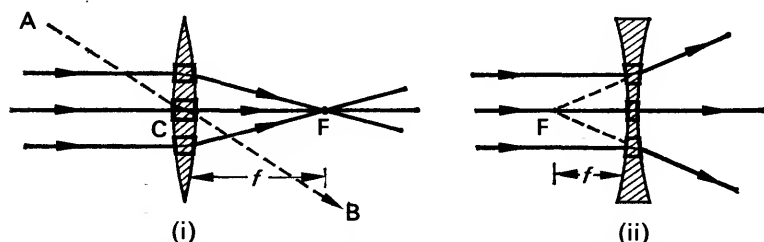


FIG. 21.11. Action of converging (convex) and diverging (concave) lenses.

imagined to be produced, the angles of the prisms can be seen to increase from zero at the middle to a small value at the edge of the lens. Now the deviation, d , of a ray of light by a small-angle prism is given by $d = (n - 1) A$, where A is the angle of the prism, see p. 457. Consequently the truncated prism corresponding to a position farther away from the middle of the lens deviates an incident ray more than those prisms nearer the middle. Thus, for the case of the converging lens, the refracted rays converge to the same point or focus F , Fig. 21.11 (i). It will be noted that a ray AC incident on the middle, C , of the lens emerges parallel to AC , since the middle acts like a rectangular piece of glass (p. 422). This fact is utilised in the drawing of images in lenses (p. 485).

Since the diverging lens is made up of truncated prisms pointing the opposite way to the converging lens, the deviation of the light is in the opposite direction, Fig. 21.11 (ii). A divergent beam is hence obtained when parallel rays are refracted by the lens.

The Signs of Focal Length, f

From Fig. 21.11 (i), it can be seen that a convex lens has a real focus; the focal length, f , of a *converging* lens is thus *positive* in sign. Since the focus of a diverging lens is virtual, the focal length of such a lens is negative in sign, Fig. 21.11 (ii). The reader must memorise the sign of f for a converging and diverging lens respectively, as this is always required in connection with lens formulae.

Relations Between Image and Object Distances for Thin Lens

We can now derive a relation between the object and image distances when a lens is used. We shall limit ourselves to the case of a *thin* lens, i.e., one whose thickness is small compared with its other dimensions, and consider narrow beams of light incident on its central portion.

Suppose a lens of refractive index n_2 is placed in a medium of refractive index n_1 , and a point object O is situated on the principal axis, Fig. 21.12.

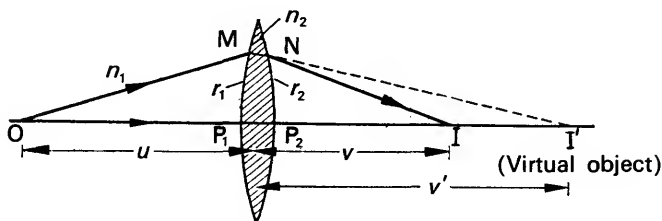


FIG. 21.12. Lens proof (exaggerated for clarity).

A ray from O through the middle of the lens passes straight through as it is normal to both lens surfaces. A ray OM from O , making a *small* angle with the principal axis, is refracted at the first surface in the direction MNI' , and then refracted again at N at the second surface so that it emerges along NI .

Refraction at first surface, MP_1 . Suppose u is the distance of the

object from the lens, i.e., $u = OP_1$, and v' is the distance of the image I' by refraction at the first surface, MP_1 , of the lens, i.e., $v' = I'P_1$. Then, since I' is situated in the medium of refractive index n_2 (I' is on the ray MN produced), we have, if $n_2 > n_1$,

$$\frac{n_2}{v'} + \frac{n_1}{u} = \frac{n_2 - n_1}{r_1} \quad (i)$$

where r_1 is the radius of the spherical surface MP_1 , see p. 472.

Refraction at second surface, NP_2 . Since MN and P_1P_2 are the incident rays on the second surface NP_2 , it follows that I' is a *virtual object* for refraction at this surface (see p. 412). Hence the object distance $I'P_2$ is negative; and as we are dealing with a thin lens, $I'P_2 = -v'$. The corresponding image distance, IP_2 or v , is positive since I is a real image. Substituting in the formula for refraction at a single spherical surface,

$$\frac{n_2}{-v'} + \frac{n_1}{v} = \frac{n_2 - n_1}{r_2} \quad (ii)$$

where r_2 is the radius of curvature of the surface NP_2 of the lens.

Lens equation. Adding (i) and (ii) to eliminate v' , we have

$$\frac{n_1}{v} + \frac{n_1}{u} = (n_2 - n_1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right);$$

and dividing throughout by n_1 ,

$$\frac{1}{v} + \frac{1}{u} = \left(\frac{n_2}{n_1} - 1 \right) \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \quad (iii)$$

Now parallel rays incident on the lens are brought to a focus. In this case, $u = \infty$ and $v = f$. From (iii),

$$\begin{aligned} \frac{1}{f} + \frac{1}{\infty} &= \left(\frac{n_2}{n_1} - 1 \right) \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \\ \therefore \frac{1}{f} &= \left(\frac{n_2}{n_1} - 1 \right) \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \end{aligned} \quad (2)$$

Substituting $\frac{1}{f}$ for the right-hand side of (iii), we obtain the important

$$\text{equation} \quad \frac{1}{v} + \frac{1}{u} = \frac{1}{f} \quad (3)$$

This is the "lens equation", and it applies equally to converging and diverging lenses if the sign convention is used (see also p. 407).

Focal Length of Lens. Small-angle Prism Method

The focal length f of a lens can also be found by using the deviation formula due to a small-angle prism. Consider a ray PQ parallel to the principal axis at a height h above it. Fig. 21.13 (i). This ray is refracted to the principal focus, and thus undergoes a small deviation through an angle d given by

$$d = \frac{h}{f} \quad (i)$$

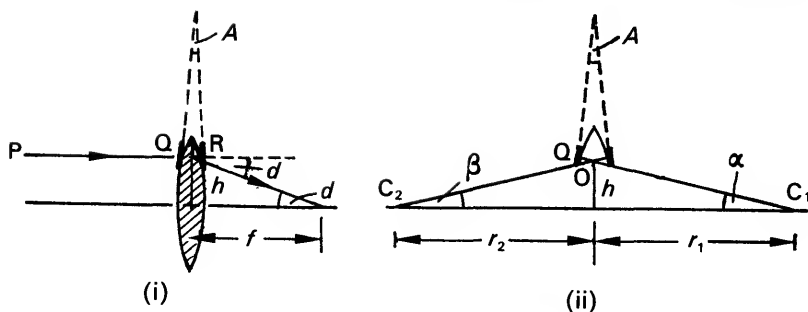


FIG. 21.13. Focal length by small angle prism.

This is the deviation through a prism of small angle A formed by the tangents at Q, R to the lens surfaces, as shown. Now for a small angle of incidence, which is the case for a thin lens and a ray close to the principal axis, $d = (n - 1) A$. See p. 457.

From (i),
$$\frac{h}{f} = (n - 1) A$$

$$\therefore \frac{1}{f} = (n - 1) \frac{A}{h} \quad \text{. (ii)}$$

The normals at Q, R pass respectively through the centres of curvatures C_1, C_2 of the lens surfaces. From the geometry, angle $ROC_1 = A = \alpha + \beta$, where α, β , are the angles with the principal axis at C_1, C_2 respectively, as shown. But $\alpha = h/r_1, \beta = h/r_2$.

$$\therefore A = \alpha + \beta = \frac{h}{r_1} + \frac{h}{r_2} \quad \text{(iii)}$$

$$\therefore \frac{A}{h} = \frac{1}{r_1} + \frac{1}{r_2}.$$

Substituting in (ii),
$$\frac{1}{f} = (n - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right).$$

Focal Length Values

Since $\frac{1}{f} = \left(\frac{n_2}{n_1} - 1 \right) \left(\frac{1}{r_1} + \frac{1}{r_2} \right)$, it follows that the focal length of a lens depends on the refractive index, n_2 , of its material, the refractive index, n_1 , of the medium in which it is placed, and the radii of curvature, r_1, r_2 , of the lens surfaces. The quantity $\frac{n_2}{n_1}$ may be termed the "relative refractive index" of the lens material; if the lens is made of glass of $n_2 = 1.5$, and it is placed in water of $n_1 = 1.33$, then the relative refractive index
$$= \frac{1.5}{1.33} = 1.13.$$

In practice, however, lenses are usually situated in air; in which case $n_1 = 1$. If the glass has a refractive index, n_2 , equal to n , the relative

refractive index, $\frac{n_2}{n_1} = \frac{n}{1} = n$. Substituting in (22), then

$$\frac{1}{f} = (n - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \quad (4)$$

Fig. 21.14 illustrates four different types of glass lenses in air, whose refractive indices, n , are each 1.5. Fig. 21.14 (i) is a biconvex lens, whose radii of curvature, r_1, r_2 , are each 10 cm. Since a spherical surface convex to a less dense medium has a positive sign (see p. 473), $r_1 = +10$ and $r_2 = +10$. Substituting in (4).

$$\therefore \frac{1}{f} = (1.5 - 1) \left(\frac{1}{(+10)} + \frac{1}{(+10)} \right) = 0.5 \times \frac{2}{10} = 0.1$$

$$\therefore f = +10 \text{ cm.}$$

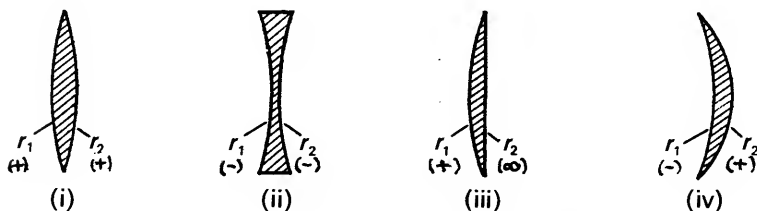


FIG. 21.14. Signs of radius of lens surface.

Fig. 21.14 (ii) is a biconcave lens in air. Since its surfaces are both concave to the less dense medium, $r_1 = -10$ and $r_2 = -10$, assuming the radii are both 10 cm. Substituting in (24),

$$\therefore \frac{1}{f} = (1.5 - 1) \left(\frac{1}{(-10)} + \frac{1}{(-10)} \right) = 0.5 \times -\frac{2}{10} = -0.1$$

$$\therefore f = -10 \text{ cm.}$$

In the case of a plano-convex lens, suppose the radius is 8 cm. Then $r_1 = +8$ and $r_2 = \infty$, Fig. 21.14 (iii). Hence $\frac{1}{f} = (1.5 - 1) \left(\frac{1}{(+8)} + \frac{1}{\infty} \right)$
 $= 0.5 \times \frac{1}{8} = \frac{1}{16}$. Thus $f = +16 \text{ cm}$.

In Fig. 21.14 (iv), suppose the radii r_1, r_2 are numerically 16 cm, 12 cm respectively. Then $r_1 = -16$, but $r_2 = +12$. Hence

$$\frac{1}{f} = (1.5 - 1) \left(\frac{1}{(-16)} + \frac{1}{(+12)} \right) = 0.5 \left(-\frac{1}{16} + \frac{1}{12} \right) = +\frac{1}{96}$$

Thus $f = +96 \text{ cm}$, confirming that the lens is a converging one.

Some Applications of the Lens Equation

The following examples should assist the reader in understanding

how to apply correctly the lens equation $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$:

1. An object is placed 12 cm from a converging lens of focal length 18 cm. Find the position of the image.

Since the lens is converging, $f = +18$ cm. The object is real, and therefore $= +12$ cm. Substituting in $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$,

$$\begin{aligned}\therefore \frac{1}{v} + \frac{1}{(+12)} &= \frac{1}{(+18)} \\ \therefore \frac{1}{v} &= \frac{1}{18} - \frac{1}{12} = -\frac{1}{36} \\ \therefore v &= -36\end{aligned}$$

Since v is negative in sign the image is *virtual*, and it is 36 cm from the lens. See Fig. 21.17 (ii).

2. A beam of light, converging to a point 10 cm behind a converging lens, is incident on the lens. Find the position of the point image if the lens has a focal length of 40 cm.

If the incident beam converges to the point O, then O is a *virtual object*, Fig. 21.15. See p. 412. Thus $u = -10$ cm. Also, $f = +40$ cm since the lens is converging. Substituting in $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$,

$$\begin{aligned}\frac{1}{v} + \frac{1}{(-10)} &= \frac{1}{(+40)} \\ \therefore \frac{1}{v} &= \frac{1}{40} + \frac{1}{10} = \frac{5}{40} \\ \therefore v &= \frac{40}{5} = 8\end{aligned}$$

Since v is positive in sign the image is *real*, and it is 8 cm from the lens. The image is I in Fig. 21.15.

3. An object is placed 6 cm in front of a diverging lens of focal length 12 cm. Find the image position.

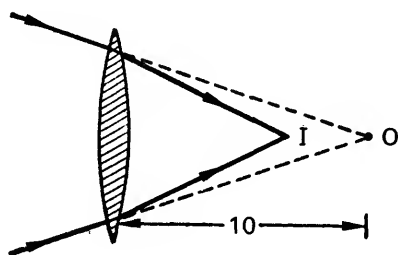


FIG. 21.15. Virtual object.

Since the lens is concave, $f = -12$ cm. The object is real, and hence $u = +6$ cm. Substituting in $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$,

$$\therefore \frac{1}{v} + \frac{1}{(+6)} = \frac{1}{(-12)}$$

$$\therefore \frac{1}{v} = -\frac{1}{12} - \frac{1}{6} = -\frac{3}{12}$$

$$\therefore v = -\frac{12}{3} = -4$$

Since v is negative in sign the image is virtual, and it is 4 cm from the lens. See Fig. 21.18 (i).

4. A converging beam of light is incident on a diverging lens of focal length 15 cm. If the beam converges to a point 3 cm behind the lens, find the position of the point image.

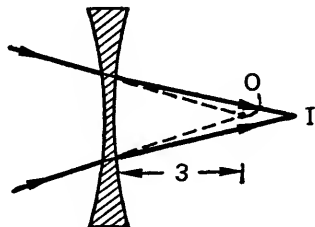


FIG. 21.16. Virtual object.

If the beam converges to the point O, then O is a virtual object, as in example 3, Fig. 21.15. Thus $u = -3$ cm. Since the lens is diverging, $f = -15$ cm. Substituting in $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$,

$$\begin{aligned}\therefore \frac{1}{v} + \frac{1}{(-3)} &= \frac{1}{(-15)} \\ \therefore \frac{1}{v} &= -\frac{1}{15} + \frac{1}{3} = \frac{4}{15} \\ \therefore v &= \frac{15}{4} = 3\frac{3}{4}\end{aligned}$$

Since v is positive in sign the point image, I, is *real*, and it is $3\frac{3}{4}$ cm from the lens, Fig. 21.16.

Images in Lenses

Converging lens. (i) When an object is a very long way from this lens, i.e., at infinity, the rays arriving at the lens from the object are parallel. Thus the image is formed at the focus of the lens, and is real and inverted.

(ii) Suppose an object OP is placed at O perpendicular to the principal axis of a thin converging lens, so that it is farther from the lens than its principal focus, Fig. 21.17 (i). A ray PC incident on the middle, C, of the lens is very slightly displaced by its refraction through the lens, as the opposite surfaces near C are parallel (see Fig. 21.11, which is an exaggerated sketch of the passage of the ray). We therefore consider that PC passes *straight through* the lens, and this is true for any ray incident on the middle of a thin lens.

A ray PL parallel to the principal axis is refracted so that it passes through the focus F. Thus the image, Q, of the top point P of the object

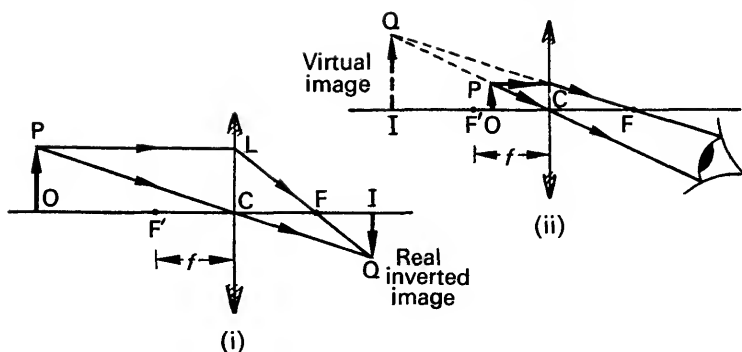


FIG. 21.17. Images in converging lenses.

is formed below the principal axis, and hence the whole image IQ is real and inverted. In making accurate drawings the lens should be represented by a straight line, as illustrated in Fig. 21.17, as we are only concerned with thin lenses and a narrow beam incident close to the principal axis.

(iii) The image formed by a converging lens is always real and inverted until the object is placed nearer the lens than its focal length, Fig. 21.17 (ii). In this case the rays from the top point P *diverge* after refraction through the lens, and hence the image Q is *virtual*. The whole image, IQ, is erect (the same way up as the object) and magnified, besides being virtual, and hence the converging lens can be used as a simple “magnifying glass” (see p. 527).

Diverging lens. In the case of a converging lens, the image is sometimes real and sometimes virtual. In a diverging lens, the image is

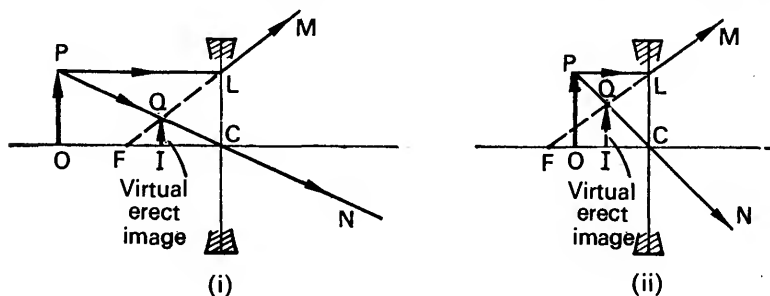


FIG. 21.18. Images in diverging lenses.

always virtual; in addition, the image is always erect and diminished. Fig. 21.18 (i), (ii) illustrate the formation of two images. A ray PL appears to diverge from the focus F after refraction through the lens, a ray PC passes straight through the middle of the lens and emerges along CN, and hence the emergent beam from P appears to diverge from Q on the same side of the lens as the object. The image IQ is thus *virtual*.

The rays entering the eye from a point on an object viewed through a lens can easily be traced. Suppose L is a converging lens, and IQ is the

image of the object OP, drawn as already explained, Fig. 21.19. If the eye E observes the top point P of the object through the lens, the cone of rays entering E are those bounded by the image Q of P and the pupil of the eye. If these rays are produced back to meet the lens L, and the

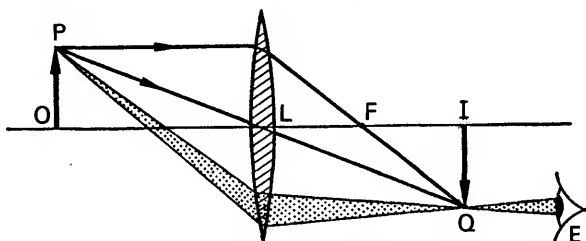


FIG. 21.19. Rays entering the eyes.

points of incidence are joined to P, the rays entering E are shown shaded in the beam. The method can be applied to trace the beam of light entering the eye from any other point on the object; the important thing to remember is to *work back from the eye*.

Another proof of $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$. We have already shown how the lens equation $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$ can be derived by considering refraction in turn at the two curved

surfaces (p. 480). A proof of the equation can also be obtained from Fig. 21.17 or Fig. 21.18, but it is not as rigid a proof as that already given on page 481.

In Fig. 21.18, triangles CQI, CPO are similar. Hence $IQ/PO = CI/CO$. Since triangles FQI, FLC are similar, $IQ/CL = FI/FC$. Now $CL = PO$. Thus the left sides of the two ratios are equal.

$$\therefore \frac{CI}{CO} = \frac{FI}{FC}$$

But $CI = -v$; $CO = +u$; $FI = FC - IC = -f - (-v) = v - f$; and $FC = -f$.

$$\begin{aligned} \therefore \frac{-v}{+u} &= \frac{v-f}{-f} \\ \therefore vf &= uv - uf \\ \therefore uf + vf &= uv \end{aligned}$$

Dividing throughout by uvf and simplifying each term,

$$\therefore \frac{1}{v} + \frac{1}{u} = \frac{1}{f}$$

The same result can be derived by considering similar triangles in Fig. 21.17, a useful exercise for the student.

Lateral Magnification

The lateral or transverse or linear magnification, m , produced by a lens is defined by

$$m = \frac{\text{height of image}}{\text{height of object}} \quad \dots \quad (5)$$

Thus $m = \frac{IQ}{OP}$ in Fig. 21.17 or Fig. 21.18. Since triangles QIC, POC are similar in either of the diagrams,

$$\frac{IQ}{OP} = \frac{CI}{CO} = \frac{v}{u},$$

where v , u are the respective image and object distances from the lens.

$$\therefore m = \frac{v}{u} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Equation (6) provides a simple formula for the magnitude of the magnification; there is no need to consider the signs of v and u in this case.

Other formulæ for magnification. Since $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, we have, by multiplying throughout by v ,

$$\begin{aligned} 1 + \frac{v}{u} &= \frac{v}{f} \\ \therefore 1 + m &= \frac{v}{f} \\ \therefore m &= \frac{v}{f} - 1 \quad . \quad . \quad . \quad . \quad . \quad (7) \end{aligned}$$

Thus if a real image is formed 25 cm from a converging lens of focal length 10 cm, the magnification, m , = $\frac{+25}{+10} - 1 = 1.5$.

By multiplying both sides of the lens equation by u , we have

$$\begin{aligned} \frac{u}{v} + 1 &= \frac{u}{f} \\ \therefore \frac{1}{m} + 1 &= \frac{u}{f} \end{aligned}$$

Object at Distance $2f$ from Converging Lens

When an object is placed at a distance of $2f$ from a convex lens, drawing shows that the real image obtained is the same size as the image

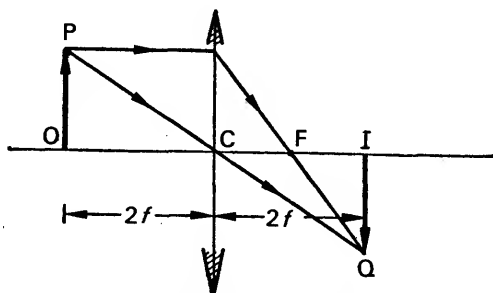


FIG. 21.20. Object and image of same size.

and is also formed at a distance $2f$ from the lens, Fig. 21.20. This result can be accurately checked by using the lens equation $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$.

Substituting $u = +2f$, and noting that the focal length, f , of a converging lens is positive, we have

$$\frac{1}{y} + \frac{1}{2f} = \frac{1}{f}$$

$$\therefore \frac{1}{v} = \frac{1}{f} - \frac{1}{2f} = \frac{1}{2f}$$

$$\therefore v = 2f = \text{image distance.}$$

$$\therefore \text{ lateral magnification, } m = \frac{v}{u} = \frac{2f}{2f} = 1,$$

showing that the image is the same size as the object.

Least Possible Distance Between Object and Real Image with Converging Lens

It is not always possible to obtain a real image on a screen, although the object and the screen may both be at a greater distance from a converging lens than its focal length. The theory below shows that the distance between an object and a screen must be equal to, or greater than, *four times the focal length* if a real image is required.

Theory. Suppose I is the real image of a point object O in a converging lens. If the image distance = x , and the distance OI = d , the object distance =

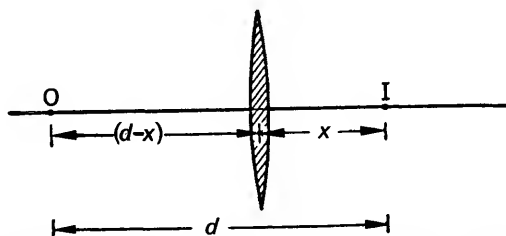


FIG. 21.21. Minimum distance between object and image.

($d - x$), Fig. 21.21. Thus $v = +x$, and $u = +(d - x)$. Substituting in the lens equation $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, in which f is positive, we have

$$\frac{1}{x} + \frac{1}{d-x} = \frac{1}{f}$$

$$\therefore \frac{d}{x(d-x)} = \frac{1}{f}$$

$$\therefore x^2 - dx + df = 0 \quad \dots \dots \dots (i)$$

For a real image, the roots of this quadratic equation for x must be real roots. Applying to (i) the condition $b^2 - 4ac > 0$ for 0 the general quadratic $ax^2 + bx + c = 0$, then

$$\begin{aligned}d^2 - 4df &> 0 \\ \therefore d^2 &> 4df \\ \therefore d &> 4f\end{aligned}$$

Thus the distance OI between the object and screen must be greater than $4f$, otherwise no image can be formed on the screen. Hence $4f$ is the minimum distance between object and screen; the latter case is illustrated by Fig. 21.20, in which $u = 2f$ and $v = 2f$. If it is difficult to obtain a real image on a screen when a converging lens is used, possible causes may be (i) the object is nearer to the lens than its focal length, Fig. 21.17 (ii), or (ii) the distance between the screen and object is less than four times the focal length of the lens.

Conjugate Points. Newton's Relation

Suppose that an object at a point O in front of a lens has its image

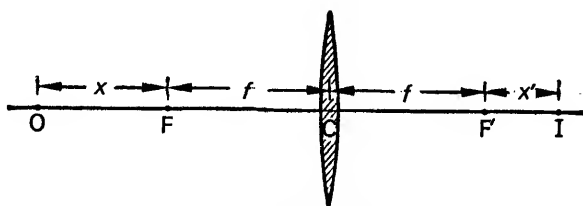


FIG. 21.22. Newton's relation.

formed at a point I. Since light rays are reversible, it follows an object placed at I will give rise to an image at O. The points O, I are thus "interchangeable", and are hence called *conjugate points* (or conjugate foci) with respect to the lens. Newton showed that conjugate points obey the relation $xx' = f^2$, where x, x' are their respective distances from the focus on the same side of the lens.

The proof of this relation can be seen by taking the case of the converging lens in Fig. 21.22, in which $OC = u = x + f$, and $CI = v = x' + f$.

Substituting in the lens equation $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$,

$$\begin{aligned}\frac{1}{x' + f} + \frac{1}{x + f} &= \frac{1}{f} \\ \therefore f(x' + x + 2f) &= (x' + f)(x + f) \\ \therefore xx' &= f^2\end{aligned}\quad (8)$$

Since $x' = f^2/x$, it follows that x' increases as x decreases. The image I thus recedes from the focus F' away from the lens when the object O approaches the lens.

The property of conjugate points stated above, namely that an object and an image at these points are interchangeable, can also be

derived from the lens equation $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$. Thus if $u = 15$ cm and $v = 10$ cm satisfies this equation, so must $u = 10$ cm and $v = 15$ cm.

Displacement of Lens when Object and Screen are Fixed

Suppose that an object O, in front of a converging lens A, gives rise to an image on a screen at I, Fig. 21.23. Since the image distance AI (v) is

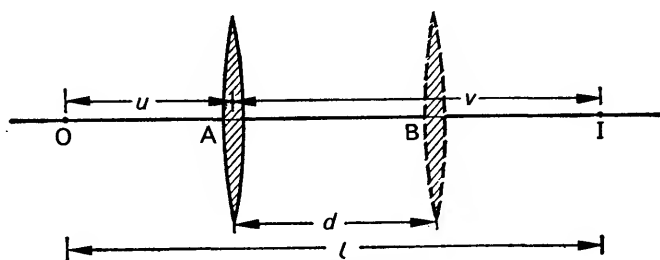


FIG. 21.23. Displacement of lens.

greater than the object distance AO (u), the image is larger than the object. If the object and the screen are kept fixed at O, I respectively, another clear image can be obtained on the screen by moving the lens from A to a position B. This time the image is smaller than the object, as the new image distance BI is less than the new object distance OB.

Since O and I are conjugate points with respect to the lens, it follows that $OB = IA$ and $IB = OA$. (If this is the case the lens equation will

be satisfied by $\frac{1}{IB} + \frac{1}{OB} = \frac{1}{f}$ and by $\frac{1}{IA} + \frac{1}{OA} = \frac{1}{f}$.) If the displacement, AB, of the lens = d , and the constant distance OI = l , then $OA + BI = l - d$. But, from above, $OA = IB$. Hence $OA = (l - d)/2$. Further, $AI = AB + BI = OA + AB = (l - d)/2 + d = (l + d)/2$.

But $u = OA$, and $v = AI$ for the lens in the position A. Substituting

for OA and AI in $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$,

$$\begin{aligned} \frac{1}{(l+d)/2} + \frac{1}{(l-d)/2} &= \frac{1}{f} \\ \therefore \frac{2}{l+d} + \frac{2}{l-d} &= \frac{1}{f} \\ \therefore \frac{4l}{l^2 - d^2} &= \frac{1}{f} \\ \therefore f &= \frac{l^2 - d^2}{4l} \end{aligned} \quad (9)$$

Thus if the displacement d of the lens, and the distance l between the object and the screen, are measured, the focal length f of the lens can be found from equation (9). This provides a very useful method of mea-

suring the focal length of a lens whose surfaces are inaccessible (for example, when the lens is in a tube), when measurements of v and u cannot be made (see p. 445).

Magnification. When the lens is in the position A, the lateral magnification m_1 of the object $= \frac{v}{u} = \frac{AI}{OA}$, Fig. 21.23.

$$\therefore \frac{h_1}{h} = \frac{AI}{AO} \quad \dots \quad (i)$$

where h_1 is the length of the image and h is the length of the object.

When the lens is in the position B, the image is smaller than the object. The lateral magnification, $m_2 = \frac{BI}{OB}$.

$$\therefore \frac{h_2}{h} = \frac{BI}{OB} \quad \dots \quad (ii)$$

where h_2 is the length of the image. But, from our previous discussion, $AI = OB$ and $OA = BI$. From (i) and (ii) it follows that, by inverting (i),

$$\begin{aligned} \frac{h}{h_1} &= \frac{h_2}{h} \\ \therefore h^2 &= h_1 h_2 \\ \therefore h &= \sqrt{h_1 h_2} \quad \dots \quad (10) \end{aligned}$$

The length, h , of an object can hence be found by measuring the lengths h_1, h_2 of the images for the two positions of the lens. This method of measuring h is most useful when the object is inaccessible, for example, when the width of a slit in a tube is required.

EXAMPLES

1. A converging lens of focal length 30 cm is 20 cm away from a diverging lens of focal length 5 cm. An object is placed 6 metres distant from the former lens (which is the nearer to it) and on the common axis of the system. Determine the position, magnification, and nature of the image formed. (O. & C.)

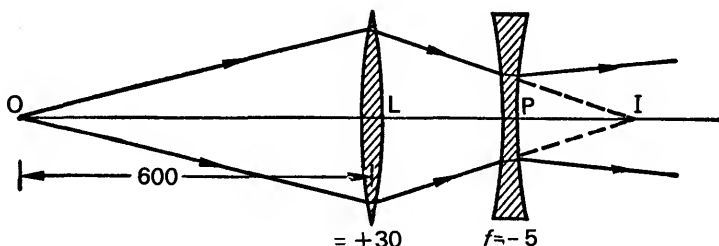


FIG. 21.24.

Suppose O is the object, Fig. 21.24.

For the converging lens,

$$u = +600 \text{ cm}, f = +30 \text{ cm}.$$

Substituting in the lens equation,

$$\therefore \frac{1}{v} + \frac{1}{(+600)} = \frac{1}{(+30)}$$

from which
$$v = \frac{600}{19} = 31\frac{11}{19} \text{ cm} = \text{LI}$$

$$\therefore \text{PI} = \text{LI} - \text{LP} = 31\frac{11}{19} - 20 = 11\frac{11}{19} \text{ cm}.$$

For the diverging lens, I is a virtual object. Thus $u = \text{PI} = -11\frac{11}{19}$. Also $f = -5$. Substituting in the lens equation, we have

$$\frac{1}{v} + \frac{1}{(-11\frac{11}{19})} = \frac{1}{(-5)}$$

from which
$$v = -8.8 \text{ cm}.$$

The image is thus virtual, and hence the rays diverge after refraction through P, as shown. The image is 8.8 cm to the left of P.

The magnification, m , is given by $m = m_1 \times m_2$, where m_1, m_2 are the magnifications produced by the converging and diverging lens respectively.

But
$$m_1 = \frac{v}{u} = 31\frac{11}{19}/600$$

and
$$m_2 = 8.8/11\frac{11}{19}$$

$$\therefore m = 31\frac{11}{19}/600 \times 8.8/11\frac{11}{19} = \frac{1}{25}$$

2. Establish a formula connecting object-distance and image-distance for a simple lens. A small object is placed at a distance of 30 cm from a converging lens of focal length 10 cm. Determine at what distances from this lens a second converging lens of focal length 40 cm must be placed in order to produce (i) an erect image, (ii) an inverted image, in each case of the same size as the object. (L.)

First part. See text.

Second part. Suppose O is the object, Fig. 21.25. The image I in the convex lens L is formed at a distance v from L given by

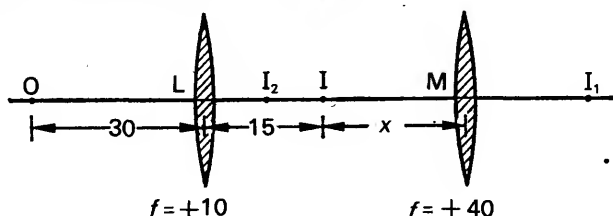


FIG. 21.25.

$$\frac{1}{v} + \frac{1}{(+30)} = \frac{1}{(+10)}$$

$v = +15.$

from which

For an erect image. Since $\frac{v}{u} = \frac{15}{30} = \frac{1}{2}$, the image at I is half the object size;

also, the image is inverted, since it is real (see Fig. 21.17 (i)). If an erect image is required, the second lens, M, must invert the image at I. Further, if the new

image, I_1 , say, is to be the same size as the object at O, the magnification produced by M of the image at I must be 2. Suppose $IM = x$ numerically; then, since the magnification $\left(\frac{v}{u}\right) = 2$, $MI_1 = 2x$. As I and I_1 are both real, we have, from

$$\frac{1}{v} + \frac{1}{u} = \frac{1}{f},$$

$$\frac{1}{(+2x)} + \frac{1}{(+x)} = \frac{1}{(+40)}$$

$$\therefore \frac{3}{2x} = \frac{1}{40}$$

from which

$$x = 60 \text{ cm.}$$

Thus M must be placed 75 cm from L for an erect image of the same size as O.

For an inverted image. Since the image at I is inverted, the image I_2 of I in M must be erect with respect to I. The lens M must thus act like a magnifying glass which produces a magnification of 2, and the image I_2 is *virtual* in this case. Suppose $IM = x$ numerically; then $I_2M = 2x$ numerically. Substituting in

$$\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$$

$$\therefore \frac{1}{(-2x)} + \frac{1}{(+x)} = \frac{1}{(+40)}$$

$$\therefore x = 20 \text{ cm.}$$

Thus M must be placed 35 cm from L for an inverted image of the same size as O.

SOME METHODS OF MEASURING FOCAL LENGTHS OF LENSES, AND THEIR RADII OF CURVATURE

Converging Lens

(1) *Plane mirror method.* In this method a plane mirror M is placed on a table, and the lens L is placed on the mirror, Fig. 21.26. A pin O is then moved along the axis of the lens until its image I is observed to coincide with O when they are both viewed from above, the method

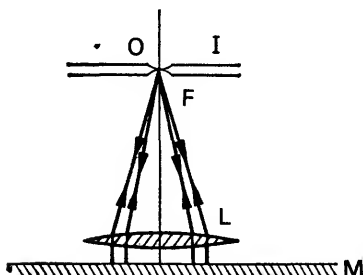


FIG. 21.26. Plane mirror method.

of no parallax being used. The distance from the pin O to the lens is then the focal length, f , of the lens, which can thus be measured.

The explanation of the method is as follows. In general, rays from O pass through the lens, are reflected from the mirror M, and then pass through the lens again to form an image at some place. When O and the image coincide in position, the rays from O incident on M must have returned along their incident path after reflection from the mirror.

This is only possible if the rays are incident *normally* on M. Consequently the rays entering the lens after reflection are all parallel, and hence the point to which they converge must be the focus, F, Fig. 21.26. It will thus be noted that the mirror provides a simple method of obtaining parallel rays incident on the lens.

(2) *Lens formula method.* In this method five or six values of u and v are obtained by using an illuminated object and a screen, or by using two pins and the method of no parallax. The focal length, f , can then be calculated from the equation $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, and the average of the values obtained. Alternatively, the values of $\frac{1}{u}$ can be plotted against $\frac{1}{v}$, and a straight line drawn through the points. When $\frac{1}{u} = 0$, $\frac{1}{v} = \frac{1}{OA} = \frac{1}{f}$, from the lens equation; thus $f = \frac{1}{OA}$, and hence can be calculated, Fig. 21.27. Since $\frac{1}{u} = \frac{1}{f}$ when $\frac{1}{v} = 0$, from the lens equation, $OB = \frac{1}{f}$. Thus f also be evaluated from $\frac{1}{OB}$.

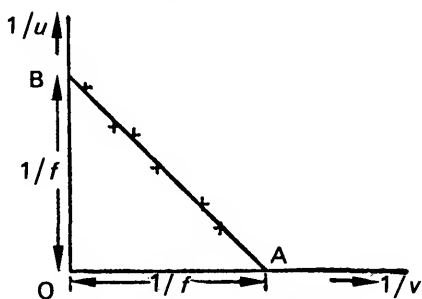


FIG. 21.27. Graph of $1/u$ against $1/v$.

(3) *Displacement method.* In this method, an illuminated object O is placed in front of the lens, A, and an image I is obtained on a screen. Keeping the object and screen fixed, the lens is then moved to a position B so that a clear image is again obtained on the screen, Fig. 21.28. From our discussion on p. 491, it follows that a magnified sharp image is obtained at I when the lens is in the position A, and a diminished sharp image when the lens is in the position B. If the displacement of the lens is d , and the distance between the object and the screen is l , the focal length, f , is given by $f = \frac{l^2 - d^2}{4l}$, from p. 491. Thus f can be calculated. The experiment can be repeated by altering the distance between the object and the screen, and the average value of f is then calculated. It should be noted that the screen must be at a distance from the object of at least four times the focal length of the lens, otherwise an image is unobtainable on the screen (p. 490).

I is farther away from L and O, since the concave lens makes the incident beam on it diverge more.

The image distance, v , from the diverging lens is CI and can be measured; v is +ve in sign as I is real. The object distance, u , from this

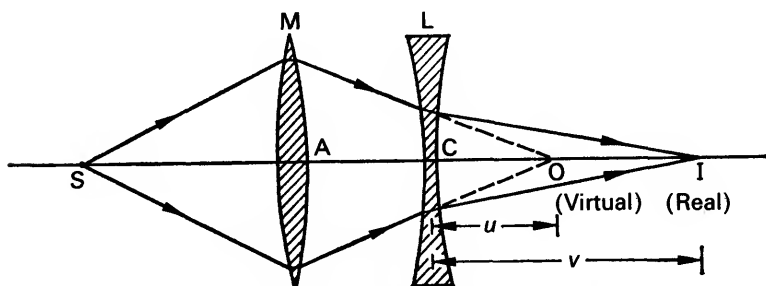


FIG. 21.30. Focal length of diverging lens.

lens = $CO = AO - AC$, and AC can be measured. The length AO is obtained by removing the lens L , leaving the converging lens, and noting the position of the real image now formed at O by the lens M . Thus $u (= CO)$ can be found; it is a -ve distance, since O is a virtual

object for the diverging lens. Substituting for u and v in $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, the focal length of the diverging lens can be calculated.

(2) *Concave mirror method.* In this method a real object is placed in front of a diverging lens, and the position of the virtual image is located with the aid of a concave mirror. An object O is placed in front of the lens L , and a concave mirror M is placed behind the lens so that a divergent beam is incident on it, Fig. 21.31. With L and M in the same position, the object O is moved until an image is obtained coincident with it in position, i.e., beside O . The distances CO , CM are then measured.

As the object and image are coincident at O , the rays must be incident *normally* on the mirror M . The rays BA , ED thus pass through the centre of curvature of M , and this is also the position of the virtual image I . The image distance, v , from the lens = $IC = IM - CM =$

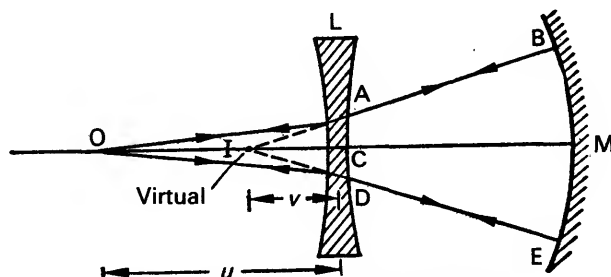


FIG. 21.31. Focal length of diverging lens.

$r = CM$, where r is the radius of curvature; CM can be measured, while r can be determined by means of a separate experiment, as described on p. 413. The object distance, u , from the lens = OC , and by substituting for u and v in the formula $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, the focal length f can be calculated. Of course, v is negative as I is a virtual image for the lens.

Measurement of Radii of Curvature of Lens Surfaces

Diverging lens. The radius of curvature of a lens concave surface A can easily be measured by moving an object O in front of it until the image by reflection at A coincides with the object. Since the rays from

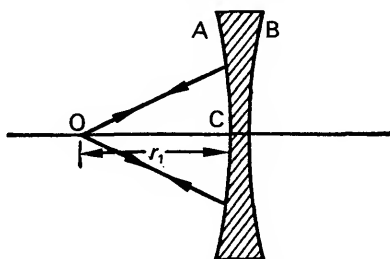


FIG. 21.32. Radius of diverging lens surface.

O are now incident normally on A , its radius of curvature, $r_1 = OC$, the distance from O to the lens, Fig. 21.32. If the radius of curvature of the surface B is required, the lens is turned round and the experiment is repeated.

Converging lens: Boys' method. Since a convex surface usually gives a virtual image, it is not an easy matter to measure the radius of curvature of such a lens surface. C. V. BOYS, however, suggested an ingenious method which is now known by his name, and is illustrated in Fig. 21.33. In Boys' method, an object O is placed in front of a converging lens, and is then moved until an image by reflection at the back surface NA is formed beside O . To make the image brighter O should be a well-illuminated object, and the lens can be floated with NA on top of mercury to provide better reflection from this surface.

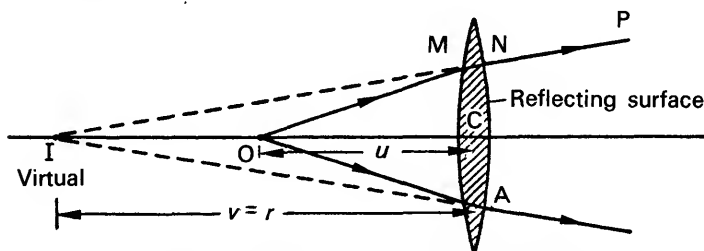


FIG. 21.33. Boys' method for radius of converging lens surface.

Since the image is coincident with O , the rays are incident normally on NA . A ray OM from O would thus pass straight through the lens

along NP after refraction at M. Further, as PN produced passes through I, I is a virtual image of O by *refraction in the lens*. On account of the latter fact, we can apply the lens equation $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, where $v = IC = r$, the radius of curvature of NA, and $u = OC$. Thus knowing OC and the focal length, f , of the lens, r can be calculated. The same method can be used to measure the radius of curvature of the surface M of the lens, in which case the lens is turned round.

Although reflection from the lens back surface is utilised, the reader should take special pains to note that Boys' method uses the formula $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$ to calculate the radius of curvature. This is a formula for the *refraction* of light through the lens.

The **refractive index** of the material of a lens can be found by measuring the radii of curvature, r_1, r_2 , of its surfaces and its focal length f . Since $\frac{1}{f} = (n - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right)$, where n is the refractive index, the latter can then be calculated.

Combined Focal Length of Two Thin Lenses in Contact

In order to diminish the colouring of the image due to dispersion when an object is viewed through a single lens, the lenses of telescopes and microscopes are made by placing two thin lenses together (see p. 515). The combined focal length, F , of the lenses can be found by considering a point object O placed on the principal axis of two *thin lenses in contact*, which have focal lengths f_1, f_2 respectively, Fig. 21.34. A ray OC from O passes through the middle, C, of both lenses undeviated.

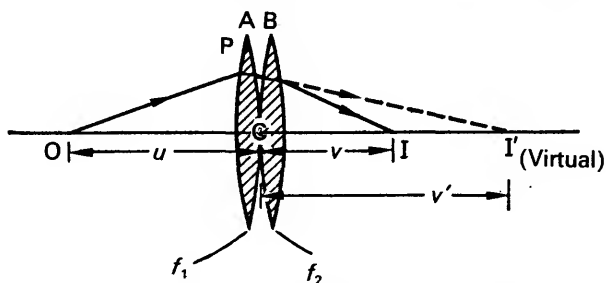


FIG. 21.34. Focal length of combined lenses.

A ray OP from O is refracted through the first lens A to intersect OC at I', which is therefore the image of O in A. If $OC = u$, $CI' = v'$,

$$\therefore \frac{1}{v'} + \frac{1}{u} = \frac{1}{f_1} \quad \dots \quad (i)$$

The beam of light incident on the second lens B converges to I', which is therefore a *virtual* object for this lens. The image is formed at I at a distance CI, or v , from the lens. Thus since the object distance

CI' is virtual, $u = -v'$ for refraction in this case. For lens B, therefore,

we have
$$\frac{1}{v} + \frac{1}{(-v')} = \frac{1}{f_2},$$

or
$$\frac{1}{v} - \frac{1}{v'} = \frac{1}{f_2} \quad \dots \quad (ii)$$

Adding (i) and (ii) to eliminate v' ,

$$\therefore \frac{1}{v} + \frac{1}{u} = \frac{1}{f_1} + \frac{1}{f_2}.$$

Since I is the image of O by refraction through both lenses,

$$\frac{1}{v} + \frac{1}{u} = \frac{1}{F},$$

where F is the focal length of the combined lenses. Hence

$$\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2} \quad \dots \quad (11)$$

This formula for F applies to any two thin lenses in contact, such as two diverging lenses, or a converging and diverging lens. When the formula is used, the signs of the focal lengths must be inserted. As an illustration, suppose that a thin converging lens of 8 cm focal length is placed in contact with a diverging lens of 12 cm focal length. Then $f_1 = +8$, and $f_2 = -12$. The combined focal length, F , is thus given by

$$\frac{1}{F} = \frac{1}{(+8)} + \frac{1}{(-12)} = \frac{1}{8} - \frac{1}{12} = +\frac{1}{24}$$

$$\therefore F = +24 \text{ cm.}$$

The positive sign shows that the combination acts like a converging lens.

The focal length of a diverging lens can be found by combining it with a converging lens of shorter focal length. The combination acts like a converging lens, as shown by the numerical example just considered, and its focal length F can be found by one of the methods described on p. 494. The diverging lens is then taken away, and the focal length f_1 of the convex lens alone is now measured. The focal length f_2 of the concave lens can then be calculated by substituting the values of F and

$$f_1 \text{ in the formula } \frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2}.$$

Refractive Index of a Small Quantity of Liquid

Besides the method given on p. 425, the refractive index of a small amount of liquid can be found by smearing it over a plane mirror and placing a converging lens on top, as shown in the exaggerated sketch of Fig. 21.35. An object O is then moved along the principal axis until the inverted image I seen looking down into the mirror is

coincident with O in position. In this case the rays which pass through the lens and liquid are incident *normally* on the mirror, and the distance from O to the lens is now the focal length, F , of the *lens and liquid* combination (see p. 494). If the experiment is repeated with the convex lens alone on the mirror, the focal length f_1 of the lens can be

measured. But $\frac{1}{F} = \frac{1}{f_1} + \frac{1}{f_2}$, where

f_2 is the focal length of the liquid lens.

Thus, knowing f_1 and F , f_2 can be calculated.

From Fig. 21.35, it can be seen that the liquid lens is a plano-concave type; its lower surface corresponds to the plane surface of the mirror, while the upper surface corresponds to the Surface S of the converging lens. If the latter has a radius of curvature r , then, from equation (4) on p. 483.

$$\begin{aligned}\frac{1}{f_2} &= (n - 1) \left(\frac{1}{r} + \frac{1}{\infty} \right). \\ \therefore \frac{1}{f_2} &= (n - 1) \frac{1}{r} \\ \therefore n - 1 &= \frac{r}{f_2} \\ \therefore n &= 1 + \frac{r}{f_2} \quad \dots \dots \dots (i)\end{aligned}$$

The radius of curvature r of the surface S of the lens can be measured by Boys' method (p. 498). Since f_2 has already been found, the refractive index n of the liquid can be calculated from (i). This method of measuring n is useful when only a small quantity of the liquid is available.

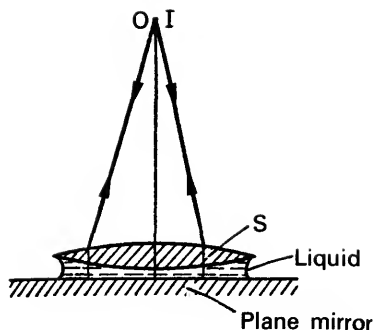


FIG. 21.35. Refractive index liquid.

EXAMPLES

1. Draw a diagram to illustrate the principle of the convex driving mirror on a motor-car. A converging lens of focal length 24 cm is placed 12 cm in front of a convex mirror. It is found that when a pin is placed 36 cm in front of the lens it coincides with its own inverted image formed by the lens and mirror. Find the focal length of the mirror.

First part. See text.

Second part. Suppose O is the position of the pin, Fig. 21.36. Since an inverted image of the pin is formed at O, the rays from O strike the convex mirror normally. Thus the image, I, in the lens of O is at the centre of curvature of the mirror.

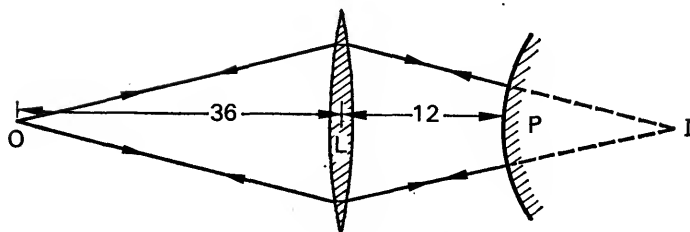


FIG. 21.36. Example.

Since $u = OL = +36$, $f = +24$, it follows from the lens equation that

$$\frac{1}{v} + \frac{1}{36} = \frac{1}{24}$$

from which

$$v = 72 \text{ cm} = LI$$

$$\therefore PI = LI - LP = 72 - 12 = 60$$

\therefore radius of curvature, r , of mirror = 60 cm.

$$\therefore \text{focal length of mirror} = \frac{r}{2} = 30 \text{ cm.}$$

2. Give an account of a method of measuring the focal length of a diverging lens, preferably without the aid of an auxiliary converging lens. A luminous object and a screen are placed on an optical bench and a converging lens is placed between them to throw a sharp image of the object on the screen; the linear magnification of the image is found to be 2.5. The lens is now moved 30 cm nearer the screen and a sharp image again formed. Calculate the focal length of the lens. (N.)

First part. A concave mirror can be used, p. 497.

Second part. If O, I are the object and screen positions respectively, and L_1, L_2 are the two positions of the lens, then $OL_1 = IL_2$, Fig. 112. See *Displacement method for focal length*, p. 495. Suppose $OL_1 = x = L_2I$.

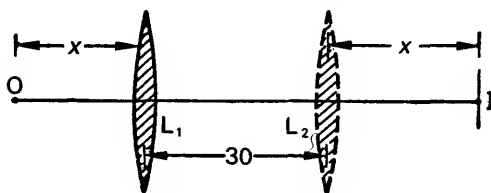


FIG. 21.37. Example.

For the lens in the position L_1 , $u = OL_1 = x$, $v = L_1I = 30 + x$.

But magnification, $m = \frac{v}{u} = 2.5$

$$\frac{30 + x}{x} = 2.5$$

$$\therefore x = 20 \text{ cm.}$$

$$\therefore u = OL_1 = 20 \text{ cm.}$$

$$v = L_1I = 30 + x = 50 \text{ cm.}$$

Substituting in $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$,

$$\therefore \frac{1}{20} + \frac{1}{50} = \frac{1}{f}$$

from which

$$f = 14\frac{2}{7} \text{ cm.}$$

3. Describe two methods for the determination of the focal length of a diverging lens. A thin equiconvex lens is placed on a horizontal plane mirror, and a pin held 20 cm vertically above the lens coincides in position with its own image. The space between the under surface of the lens and the mirror is filled with water (refractive index 1.33) and then, to coincide with its image as before, the pin has to be raised until its distance from the lens is 27.5 cm. Find the radius of curvature of the surfaces of the lens. (N.)

First part. See text.

Second part. The focal length, f_1 , of the lens = 20 cm, and the focal length, F , of the water and glass lens combination = 27.5 cm. See p. 501 and Fig. 21.35. The focal length, f , of the water lens is given by

$$\frac{1}{F} = \frac{1}{f} + \frac{1}{f_1}$$

$$\therefore \frac{1}{(+27.5)} = \frac{1}{f} + \frac{1}{(+20)}$$

Solving, $\therefore \frac{1}{f} = \frac{1}{27.5} - \frac{1}{20} = -\frac{3}{220}$

the minus showing that the water lens is a diverging lens.

But $\frac{1}{f} = (n - 1)\frac{1}{r}$,

where $n = 1.33$, and r = radius of the curved face of the lens.

$$\therefore -\frac{3}{220} = (1.33 - 1)\frac{1}{r}$$

$$\therefore r = -24.2 \text{ cm.}$$

The glass lens is equiconvex, and hence the radii of its surfaces are the same.

4. Derive an expression for the equivalent focal length of a system of two thin lenses of focal lengths f_1 and f_2 in contact. Two equiconvex lenses of focal length 20 cm are placed in contact and the space between them filled with water. Find the focal length of the combination ($n_g = 3/2$, $n_w = 4/3$). (L.)

First part. See text.

Second part. Since the lenses are equiconvex, the radii of curvature, r , of their surfaces are equal. Now

$$\frac{1}{f} = (n - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right)$$

$$\therefore \frac{1}{20} = \left(\frac{3}{2} - 1 \right) \left(\frac{1}{r} + \frac{1}{r} \right)$$

$$\therefore r = 20 \text{ cm.}$$

The water between the lenses forms an equiconcave lens of refractive index $4/3$ and radii 20 cm. Its focal length f_1 is thus given by

$$\begin{aligned}\frac{1}{f_1} &= (n - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \\ &= \left(\frac{4}{3} - 1 \right) \left(\frac{1}{-20} + \frac{1}{-20} \right) \\ \therefore f_1 &= -30 \text{ cm.}\end{aligned}$$

The focal length, F , of the combination is given by

$$\frac{1}{F} = \frac{1}{f} + \frac{1}{f_1} + \frac{1}{f} = \frac{2}{f} + \frac{1}{f_1}$$

where f is the focal length of a glass lens.

$$\begin{aligned}\therefore \frac{1}{F} &= \frac{2}{(+20)} + \frac{1}{(-30)} = \frac{2}{30} \\ \therefore F &= \frac{30}{2} = 15 \text{ cm.}\end{aligned}$$

EXERCISES 21

Refraction at a Single Curved Surface

1. A solid glass sphere has a radius of 10 cm and a refractive index of 1.5. Find the position from the centre, and nature, of the image of an object (i) 20 cm, (ii) 40 cm from the centre due to refraction at the nearest part of the sphere.

2. Obtain a formula connecting the distances of the object and the image from a spherical refracting surface. A transparent sphere of refractive index $4/3$ has a radius of 12 cm. Find the positions of the image of a small object inside it 4 cm from the centre, when it is viewed first on one side and then on the other side of the sphere, in the direction of the line joining the centre to the object.

3. Viewed normally through its flat surface, the greatest thickness of a plano-convex lens appears to be 2.435 cm, and through its curved surface 2.910 cm. Actually it is 3.665 cm. Find (a) the refractive index of the glass, (b) the radius of curvature of the convex surface. Do you consider this is a satisfactory method of finding the radius of curvature? (*N.*)

4. A large glass sphere is placed immediately behind a small hole in an opaque screen, and a small filament lamp is placed at such a distance u in front of the hole that its image falls within the sphere, and at a distance v behind the hole. (a) Sketch the course taken by the light rays in the formation of this image. (b) Derive a formula connecting the quantities u and v with the refractive index n of the glass. (c) If $n = 1.5$ and $u = 4r$, where r is the sphere's radius, find the image position due to refraction at the nearest part of the glass surface only.

Refraction Through Lenses

5. An object is placed (i) 12 cm, (ii) 4 cm from a converging (convex) lens of focal length 6 cm. Calculate the image position and the magnification in each case, and draw sketches illustrating the formation of the image.

6. What do you know about the image obtained with a diverging (concave) lens? The image of a real object in a diverging lens of focal length 10 cm is formed 4 cm from the lens. Find the object distance and the magnification. Draw a sketch to illustrate the formation of the image.

7. The image obtained with a converging lens is erect and three times the length of the object. The focal length of the lens is 20 cm. Calculate the object and image distances.

8. A beam of light converges to a point 9 cm behind (i) a converging lens of focal length 12 cm, (ii) a diverging lens of focal length 15 cm. Find the image position in each case, and draw sketches illustrating them.

9. (i) The surfaces of a biconvex lens are 8 cm and 12 cm radius of curvature. If the refractive index of the glass is 1.5, calculate the focal length. (ii) The curved surface of a plano-concave lens is 10 cm radius of curvature, and the refractive index of the glass is 1.6. Calculate the focal length.

10. Describe an experiment to obtain an accurate value for the focal length of a thin converging lens.

A thin converging lens is fixed inside a tube AB . A sharp image of an illuminated object is formed on a screen when the end A of the tube is 90.0 cm from the screen and again when it is 140.0 cm from the screen. If the distance between object and screen is 250 cm in each case, how far is the lens from A ? (L .)

11. Why is a sign convention used in geometrical optics?

A thin equiconvex lens of glass of refractive index 1.50 whose surfaces have a radius of curvature of 24.0 cm is placed on a horizontal plane mirror. When the space between the lens and mirror is filled with a liquid, a pin held 40.0 cm vertically above the lens is found to coincide with its own image. Calculate the refractive index of the liquid. (N .)

13. State clearly the sign convention you employ in optics. Light from a point O in air on the axis of a simple spherical interface between air and glass is refracted so as to form an image at I . Derive a formula connecting the distances of O and I from the pole when I is (a) real, (b) virtual. Show that the two formulae can be reduced to a single formula by use of your sign convention.

Calculate the focal length in air of a thin converging meniscus lens with surfaces having radii of curvature 16.0 cm and 24.0 cm, the refractive index of the glass being 1.60.

Indicate briefly a method for measuring this focal length. (L .)

12. What do you understand by a *virtual object* in optics? Describe a direct method, not involving any calculation, of finding (a) the focal length of a double concave lens, and (b) the radius of curvature of one of its faces. You may use other lenses and mirrors if you wish.

A converging lens of 20 cm focal length is arranged coaxially with a diverging lens of focal length 8.0 cm. A point object lies on the same side as the converging lens and very far away on the axis. What is the smallest possible distance between the lenses if the combination is to form a real image of the object?

If the lenses are placed 6.0 cm apart, what is the position and nature of the final image of this distant object? Draw a diagram showing the passage of a wide beam of light through the system in this case. (C .)

14. Explain in detail how, with the aid of a pin and a plane mirror, you would determine the focal length of a thin biconvex lens.

Having found the focal length of this lens, explain how you would find the radius of curvature of one of its faces by Boys' method. Discuss whether or not this method can be used to find the radii of curvature of the faces of a thin converging meniscus lens.

The radii of curvature of the faces of a thin converging meniscus lens of material of refractive index $\frac{3}{2}$ are 10 cm and 20 cm. What is the focal length of the lens (a) in air, (b) when completely immersed in water of refractive index $\frac{4}{3}$? (N.)

15. A thin planoconvex lens is made of glass of refractive index 1.5. When an object is set up 10 cm from the lens, a virtual image ten times its size is formed. What is (a) the focal length of the lens, (b) the radius of curvature of the lens surface?

If the lens is floated on mercury with the curved side downwards and a luminous object placed vertically above it, how far must the object be from the lens in order that it may coincide with the image produced by reflection in the curved surface? (L.)

16. Deduce an expression for the focal length of a lens in terms of u and v , the object and image distance from the lens.

A lens is set up and produces an image of a luminous point source on a screen 25 cm away. *If the aperture of the lens is small*, where must the screen be placed to receive the image when a parallel slab of glass 6 cm thick is placed at right angles to the axis of the lens and between the lens and the screen, if the refractive index of the glass is 1.6? Deduce any formula you use. (L.)

17. Derive an expression for the focal length of a lens in terms of the radii of curvature of its faces and its refractive index.

Find the condition that the distance between the object and image is a minimum, and explain how you would verify your result experimentally. (C.)

18. Give an account of a method of finding the focal length of a thin concave lens using an auxiliary convex lens which is not placed in contact with it.

A thin equiconvex lens of refractive index 1.50 is placed on a horizontal plane mirror, and a pin fixed 15.0 cm above the lens is found to coincide in position with its own image. The space between the lens and the mirror is now filled with a liquid and the distance of the pin above the lens when the image and object coincide is increased to 27.0 cm. Find the refractive index of the liquid. (N.)

19. A thin converging lens is mounted coaxially inside a short cylindrical tube whose ends are closed by means of thin windows of parallel-sided glass. Explain the *principles* of two methods by which the focal length of the lens could be determined.

When a thin biconvex lens is placed on a horizontal mirror, a pin placed 14.0 cm above the lens on the axis is found to coincide with its own image. When a little water of refractive index 1.33 is inserted between lens and mirror the self-conjugate positions of the pin are respectively 17.2 cm and 26.2 cm above the lens, with first one and then the other face of the lens in contact with the water.

Deduce what information you can about the lens from these data. (L.)

20. Describe in detail how you would determine the focal length of a diverging lens with the help of (a) a converging lens, (b) a concave mirror.

A converging lens of 6 cm focal length is mounted at a distance of 10 cm from a screen placed at right angles to the axis of the lens. A diverging lens of 12 cm focal length is then placed coaxially between the converging lens and the screen so that an image of an object 24 cm from the converging lens is focused on the screen. What is the distance between the two lenses? Before commencing the calculation state the sign convention you will employ. (N.)

21. (a) Find an expression for the focal length of a combination of two thin lenses in contact. (b) A symmetrical convex glass lens, the radii of curvature of which are 3 cm, is situated just below the surface of a tank of water which is 40 cm deep. An illuminated scratch on the bottom of the tank is viewed vertically downwards through the lens and the water. Where is the image, and where should the eye of the observer be placed in order to see it? The refractive indices of glass and water may be taken as $\frac{3}{2}$ and $\frac{4}{3}$ respectively. (O. & C).

22. Find the relation between the focal lengths of two thin lenses in contact and the focal length of the combination.

The curved face of a planoconvex lens ($n = 1.5$) is placed in contact with a plane mirror. An object at 20 cm distance coincides with the image produced by the lens and reflection by the mirror. A film of liquid is now placed between the lens and the mirror and the coincident object and image are at 100 cm distance. What is the index of refraction of the liquid? (L.)

23. Describe an optical method of finding the radius of curvature of a surface of a thin convex lens.

An object is placed on the axis of a thin planoconvex lens, and is adjusted so that it coincides with its own image formed by light which has been refracted into the lens at its first surface, internally reflected at the second surface, and refracted out again at the first surface. It is found that the distance of the object from the lens is 20.5 cm when the convex surface faces the object, and 7.9 cm when the plane surface faces the object. Calculate (a) the focal length of the lens, (b) the radius of curvature of the surface, (c) the refractive index of the glass. (C).

24. Define *focal length*, *conjugate foci*, *real image*. Obtain an expression for the transverse magnification produced by a thin converging lens.

Light from an object passes through a thin converging lens, focal length 20 cm, placed 24 cm from the object and then through a thin diverging lens, focal length 50 cm, forming a real image 62.5 cm from the diverging lens. Find (a) the position of the image due to the first lens, (b) the distance between the lenses, (c) the magnification of the final image. (L.)

25. Describe how you would determine the focal length of a diverging lens if you were provided with a converging lens (a) of shorter focal length, (b) of longer focal length.

An illuminated object is placed at right angles to the axis of a converging lens, of focal length 15 cm, and 22.5 cm from it. On the other side of the converging lens, and coaxial with it, is placed a diverging lens of focal length 30 cm. Find the position of the final image (a) when the lenses are 15 cm apart and a plane mirror is placed perpendicular to the axis 40 cm beyond the diverging lens, (b) when the mirror is removed and the lenses are 35 cm apart. (N.)

chapter twenty-two

Defects of vision Defects of lenses

DEFECTS OF VISION

THERE are numerous defects of vision, each necessitating the use of a different kind of spectacles. The use of convex lenses in spectacles was fairly widespread by 1300, but concave lenses were not in common use until about 1550, and were then highly valued. We propose to discuss briefly the essential optical principles of some of the main defects of vision and their "correction", and as a necessary preliminary we must mention certain topics connected with the eye itself.

Far and Near Points of Eye. Accommodation

An account of the essential features of the eye was given in Chapter 15 on p. 389, and it was mentioned there that the image formed by the eye-lens L must appear on the retina R , the light-sensitive screen of the eye, in order to be clearly seen, Fig. 22.1. The ciliary muscles enable the eye to focus objects at different distances from it, a property of the eye known as its power of *accommodation*. The most distant point it can focus (the "*far point*") is at infinity for a normal eye; and as the ciliary muscles are then completely relaxed, the eye is said to be "unaccommodated", or "at rest". In this case parallel rays entering the eye are focused on the retina, Fig. 22.1 (i).

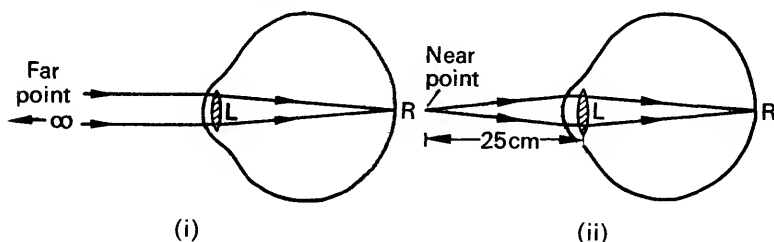


FIG. 22.1. Normal eye.

On the other hand, an object is seen in great detail when it is placed as near the eye as it can be while remaining focused; this distance from the eye known as its *least distance of distinct vision*. The point at this distance from the eye is called its *near point*, and the distance is about 25 cm for a normal eye, Fig. 22.1 (ii). The eye is said to be "fully accom-

modated" when viewing an object at its near point, as the ciliary muscles are then fully strained.

Short Sight

If the focal length of the eye is too short, owing to the eye-ball being too long, parallel rays will be brought to a focus at a point D in front of the retina, Fig. 22.2 (i). In this case the far point of the eye is not at infinity, but at a point P nearer to the eye, and the defect of vision is known as *short sight*, or *myopia*.

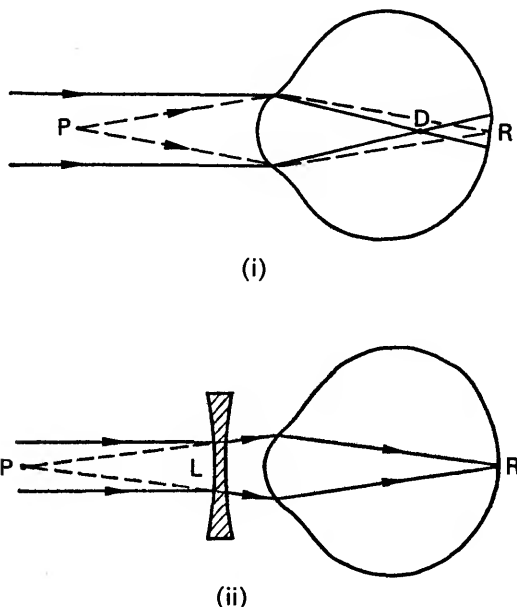


FIG. 22.2. Short sight and its correction.

A suitable *diverging* lens, L, is required to correct for short sight, Fig. 22.2 (ii). Parallel rays refracted through L are now made divergent, and if they appear to come from the far point P of the eye, the rays are brought to a focus on the retina R. From Fig. 22.2 (ii), it can be seen that the focal length of the required lens is equal to PL, which is practically equal to the distance of the far point from the eye.

Long sight

If a person's far point is normal, i.e. at infinity, but his near point is farther from the eye than the normal least distance of distinct vision, 25 cm, the person is said to be "long-sighted". In Fig. 22.3 (i), X is the near point of a person suffering from long sight, due to a short eyeball, for example. Rays from X are brought to a focus on the retina R; where-

as rays from the normal near point A, 25 cm from the eye, are brought to a focus at B behind the retina.

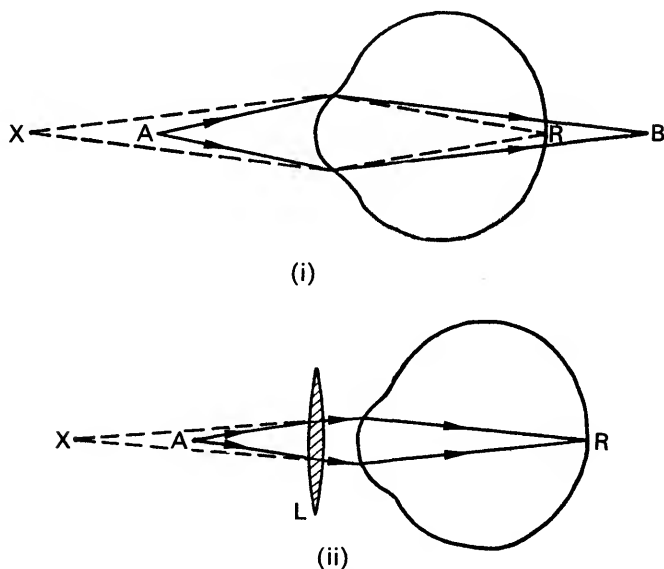


FIG. 22.3. Long sight and its correction for near point.

A suitable *converging* lens, L, is required to correct for this defect of vision, Fig. 22.3 (ii). Rays from A then appear to come from X after refraction through L, and an image is thus now formed on the retina. It can be seen that X is the virtual image of A in the lens L. Thus if $XL = 50$ cm, and $AL = 25$ cm, the focal length of L is given from the lens formula by

$$\begin{aligned}\frac{1}{(-50)} + \frac{1}{(+25)} &= \frac{1}{f} \\ \therefore \frac{1}{50} &= \frac{1}{f} \\ \therefore f &= 50 \text{ cm.}\end{aligned}$$

Correction for far point. Presbyopia

Some people are able to focus only those beams of light which converge to a point behind the retina, in which case the far point is virtual. A parallel beam of light is then brought to a focus behind the retina, R, Fig. 22.4 (i). This defect of vision is corrected by using a suitable converging lens L, Fig. 22.4 (ii).

Short sight and long sight are defects of refraction. A person with these defects of vision still has the power of accommodation. *Presbyopia*, however, is a defect of vision where the eyelens becomes inelastic through old age. Thus the eye is unable to accommodate. *Bifocals*, to

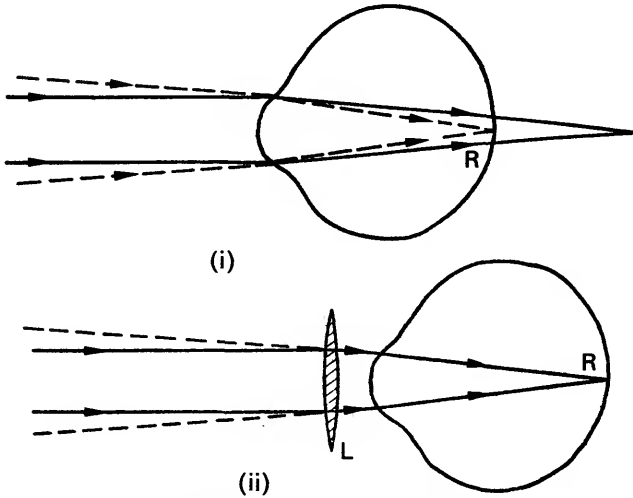


FIG. 22.4. Long sight, and its correction for far point.

correct for both near and far points, may be required. Fig. 22.5 illustrates a bifocal lens, with a positive reading section below and a negative section above for far vision.

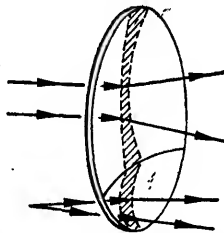
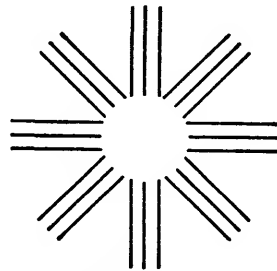


FIG. 22.5. Bifocal lens. Correction for near and far vision.

Astigmatism

If the cornea, the refracting surface in front of the eye (p. 389), has widely varying curvatures in different planes, the rays from an object in one plane will be brought to a focus by the eye at a different place from rays in another plane. In this case the lines in one direction in Fig. 22.6 will be sharply focused compared with the other lines. A strain is thus imposed on the eye when viewing an object, and this defect of vision is called *astigmatism*. It is corrected by using a *cylindrical lens*, whose curvature compensates for the curvature of the cornea in the particular astigmatic plane.

FIG. 22.6.
Test of astigmatism.

EXAMPLES

1. Draw diagrams to illustrate *long sight* and *short sight*. Draw also diagrams showing the correction of these defects by suitable lenses. A person can focus objects only when they lie between 50 cm and 300 cm from his eyes. What spectacles should he use (a) to increase his maximum distance of distinct vision to infinity, (b) to reduce his least distance of distinct vision to 25 cm? Find this range of distinct vision using each pair. (N)

First part. See text.

Second part. (a) To increase the maximum distance of distinct vision from 300 cm to infinity, the person requires a *diverging* lens. See Fig. 22.2 (ii). Assuming the lens is close to his eye, the focal length $PL = 300$ cm as P is 300 cm from the eye. One limit of the range of distinct vision is now infinity. The other limit is the object (u) corresponding to an image distance (v) of 50 cm from the lens, as the person can see distinctly things 50 cm from his eyes. In this case, then, $v = -50$ cm, $f = -300$ cm. Substituting in the lens equation, we have

$$\frac{1}{(-50)} + \frac{1}{u} = \frac{1}{(-300)}$$

from which

$$u = 60 \text{ cm.}$$

The range of distinct vision is thus from 60 cm to infinity.

(b) To reduce his least distance of distinct vision from 50 cm to 25 cm, the person requires a *converging* lens. See Fig. 22.3 (ii). In this case, assuming the lens is close to the eye, $u = +25$ cm, $v = -50$ cm, as the image must be formed 50 cm from the eye on the same side as the object, making the image virtual. The focal length of the lens is thus given by

$$\frac{1}{(-50)} + \frac{1}{(+25)} = \frac{1}{f}$$

from which

$$f = +50 \text{ cm.}$$

Objects placed at the focus of this lens appear to come from infinity. The maximum distance of distinct vision, u , is given by substituting $v = -300$ cm and $f = +50$ cm in the lens formula.

Thus

$$\frac{1}{(-300)} + \frac{1}{u} = \frac{1}{(+50)}$$

from which

$$u = \frac{300}{7} = 42\frac{2}{7} \text{ cm.}$$

The range of distinct vision is thus from 25 to $42\frac{2}{7}$ cm.

2. Explain what is meant by the magnifying power of an optical instrument, considering the cases of microscope and telescope. A thin converging lens of focal length 5 cm is laid on a map situated 60.5 cm below the eye of an observer whose least distance of distinct vision is 24.5 cm. Describe what is seen (a) then, and when the lens is raised (b) 5 cm, (c) $5\frac{1}{2}$ cm, (d) 6 cm above the map. (L.)

First part. See later text, p. 526.

Second part. (a) When the lens is on top of the map, it acts as a thin piece of glass which very slightly raises the map to the observer (p. 427). The map appears unaltered in size and is the same way up.

(b) When the map is 5 cm from the lens, i.e., at its focal plane, the map is the same way up but now appears bigger. In this case the image is at infinity.

(c) When the map is $5\frac{1}{2}$ cm from the lens, the image is inverted. Its distance v is given by $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, from which

$$\frac{1}{v} + \frac{1}{(+5\frac{1}{2})} = \frac{1}{(+5)}$$

or $v = 55$ cm.

The image is thus $60.5 - (55 + 5\frac{1}{2})$ cm from the observer, i.e., the image is formed at the eye. A blurred image is seen.

(d) When the map is 6 cm from the lens, the image is inverted and its distance given by

$$\frac{1}{v} + \frac{1}{(+6)} = \frac{1}{(+5)}$$

$\therefore v = 30$ cm.

The image is thus $60.5 - (30 + 5.5)$ cm, or 25 cm, from the eye of the observer. The inverted map is thus seen clearly. Since $v/u = 30/6 = 5$, the inverted map is magnified five times.

DEFECTS OF LENSES

We have just considered defects inherent in the eye; we have now to consider defects of an image produced by a lens, which is quite a different matter.

Chromatic Aberration

The image of an object formed by a single lens is distorted from a variety of causes. The main defect of the lens is the colouring of the image it produces, which is known as *chromatic aberration*.

Experiment shows that if a parallel beam of white light is incident on a converging lens, the red rays in the light are brought to a focus R, and the blue rays are brought to slightly nearer focus at B, Fig. 22.7.

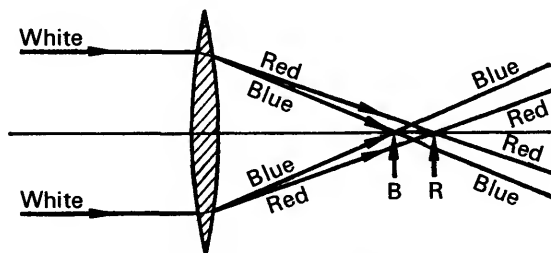


FIG. 22.7. Chromatic aberration.

Thus a single lens produces coloured images of an object which are at slightly different positions. Because he did not know how to eliminate the chromatic aberration Newton decided to abandon the use of lenses for large telescopes (p. 533).

It has already been noted that the refractive index, n , of a material varies with the colour of the light (p. 458). Thus since $1/f = (n - 1) \times \left(\frac{1}{r_1} + \frac{1}{r_2}\right)$, where f is the focal length of a lens, n is the refractive index of the material, and r_1, r_2 are the radii of curvature of the surfaces, it follows that f has different values for different colours. For example, the focal length of a lens for blue light, f_b , is given by $\frac{1}{f_b} = (n_b - 1) \times \left(\frac{1}{r_1} + \frac{1}{r_2}\right)$, where n_b is the refractive index for blue light. Suppose the lens in Fig. 22.7 is made of crown glass, for which $n_b = 1.523$, and suppose r_1, r_2 are 15 and 12 cm respectively. Then

$$\frac{1}{f_b} = (1.523 - 1) \left(\frac{1}{(+15)} + \frac{1}{(+12)} \right)$$

from which $f_b = 12.86$ cm.

For the same glass, the refractive index n_r for red light = 1.513. The focal length f_r for red light is hence given by

$$\begin{aligned} \frac{1}{f_r} &= (n_r - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \\ &= (1.513 - 1) \left(\frac{1}{(+15)} + \frac{1}{(+12)} \right), \end{aligned}$$

from which $f_r = 13.00$ cm.

The separation BR of the two foci is thus given by

$$f_r - f_b = 13.00 - 12.86 = 0.14 \text{ cm.}$$

The ratio of the focal lengths for the two colours = $\frac{f_r}{f_b} = \frac{n_b - 1}{n_r - 1}$,

since $\frac{1}{f_b} = (n_b - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right)$ and $\frac{1}{f_r} = (n_r - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right)$.

Achromatic Combination of Lenses

A converging lens deviates an incident ray such as AB towards its principal axis, Fig. 22.8 (i). A diverging lens, however, deviates a ray PQ away from its principal axis, Fig. 22.8 (ii). The dispersion between two colours produced by a converging lens can thus be annulled by

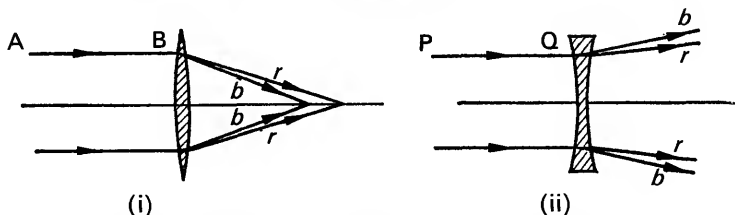


FIG. 22.8. Dispersion produced by convex and concave lens.

placing a suitable diverging lens beside it. Except for the mathematics, the making of *achromatic lenses* is analogous to the case of achromatic prisms, discussed on p. 459. There it was shown that two prisms of different material, with angles pointing in opposite directions, can act as an achromatic combination. Fig. 22.9 illustrates an achromatic lens combination, known as an *achromatic doublet*. The biconvex lens is made of crown glass, while the concave lens is made of flint glass and is a plano-concave lens. So that the lenses can be cemented together with Canada balsam, the radius of curvature of the curved surface of the plano-concave lens is made numerically the same as that of one surface of the convex lens. The achromatic combination acts as a convex lens when used as the objective lens in a telescope or microscope.

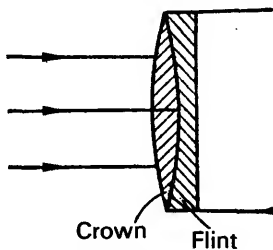


FIG. 22.9.
Achromatic doublet in
telescopic objective.

It should be noted that chromatic aberration would occur if the diverging and converging lenses were made of the *same* material, as the two lenses together would then constitute a single thick lens of one material.

Condition for an Achromatic Combination

Achromatic lenses were first made about 1729, years after Newton had considered they were impossible to construct. The necessary condition for an achromatic combination of lenses is derived on p. 516. Here we shall accept the result, which states: *Two lenses form an achromatic doublet for two colours, if the ratio of their focal lengths is numerically equal to the ratio of the corresponding dispersive powers* (p. 459) *of their materials*. Hence, since f_1, f_2 are of opposite signs because one lens must be convex and the other concave,

$$\frac{f_1}{f_2} = -\frac{\omega_1}{\omega_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where f_1, f_2 are the mean focal lengths and ω_1, ω_2 are the respective dispersive powers.

If a lens combination of focal length F is required, then f_1, f_2 must satisfy in addition the relation

$$\frac{1}{f_1} + \frac{1}{f_2} = \frac{1}{F} \quad . \quad . \quad . \quad . \quad . \quad . \quad (i)$$

Knowing ω_1, ω_2 , and F , the magnitudes of f_1, f_2 can be found by solving the equations (i) and (1).

There still remains, however, the practical matter of fitting the surfaces of the two lenses together, as the lenses must be in good optical contact to function efficiently. Suppose the flint glass lens has one surface plane, as shown in Fig. 22.9, and has the focal length f_1 . Then

$$\frac{1}{f_1} = (n - 1) \left(\frac{1}{r_1} + \frac{1}{\infty} \right) = (n - 1) \frac{1}{r_1},$$

where n is the mean refractive index of flint glass and r_1 is the radius

of curvature of the lens. Since f_1 and n are known, r_1 can be calculated. The focal length f_2 of the crown glass convex lens is given by

$$\frac{1}{f_2} = (n' - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right)$$

where n' is the mean refractive index of crown glass and r_2 is the radius of curvature of the other lens surface. Knowing f_2 , n' , and r_1 , the magnitude of r_2 can be calculated. In this way the lens manufacturer knows what the radii of curvature of the crown and flint glass lenses must be to form an achromatic doublet of a specified focal length F .

Condition for achromatic lenses. Since $\frac{1}{f} = (n - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right)$ with the usual notation,

$$\frac{1}{f_b} = (n_b - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right)$$

and

$$\frac{1}{f_r} = (n_r - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right)$$

$$\therefore \frac{1}{f_b} - \frac{1}{f_r} = (n_b - n_r) \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \quad \dots \quad (i)$$

Now the magnitudes of $\frac{1}{f_b}$ and $\frac{1}{f_r}$ are very close to each other since f_b is nearly equal to f_r . Thus, using the calculus notation, $\frac{1}{f_b} - \frac{1}{f_r} = \delta \left(\frac{1}{f} \right)$; the latter represents the small change in $\frac{1}{f}$ when blue, and then red, rays are incident on the lens. From (i),

$$\delta \left(\frac{1}{f} \right) = (n_b - n_r) \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \quad \dots \quad (ii)$$

But

$$\frac{1}{f} = (n - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right) \quad \dots \quad (iii)$$

where f is the focal length of the lens when yellow light is incident on the lens, and n is the refractive index for yellow light. Dividing (ii) by (iii) and simplifying for $\delta \left(\frac{1}{f} \right)$, we obtain

$$\delta \left(\frac{1}{f} \right) = \frac{n_b - n_r}{n - 1} \cdot \frac{1}{f}$$

Now the dispersive power, ω , of a material is defined by the relation

$$\omega = \frac{n_b - n_r}{n - 1} \quad (\text{see p. 459.})$$

$$\therefore \delta \left(\frac{1}{f} \right) = \frac{\omega}{f} \quad \dots \quad (iv)$$

Combined lenses. Suppose f_1, f_2 are the respective focal lengths of two thin lenses in contact, ω_1, ω_2 are the corresponding dispersive powers of their materials, and F is the combined focal length. If the combination is achromatic for blue and red light, the focal length F_b for blue light is the same as the focal length F_r for red light, i.e., $F_b = F_r$.

Eye-pieces. Huygens' Eye-piece

The eye-piece of a telescope should be designed with a view to reducing chromatic and spherical aberration; in practice this can most conveniently be done by using *two* lenses as the eye-piece. We have already shown that such lenses, made of the same material, are achromatic if their distance apart is equal to the average of their focal lengths. Further, the lenses reduce spherical aberration if their distance apart is equal to the difference between their focal lengths.

Huygens designed an eye-piece consisting of two plano-convex lenses; one lens, F, had three times the focal length of the other, E, Fig. 22.11 (i). The lens F pointing to the telescope objective is known as

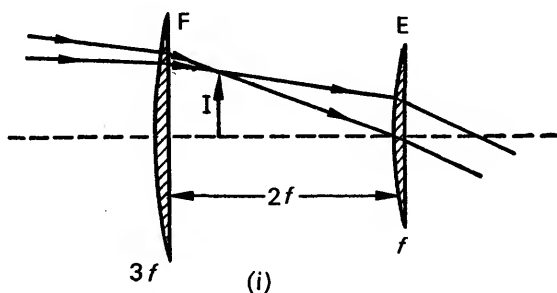


FIG. 22.11 (i). Huygens' eye-piece.

the *field lens*, while the lens E close to the eye is known as the *eye lens*, and F and E are at a distance $2f$ apart, where f is the focal length of E. Since $3f$ is the focal length of F, it follows from above that the eye-piece eliminates chromatic aberration and reduces spherical aberration.

Since the image formed by the objective of a telescope is at a distance equal to, or less than, the focal length of the eye lens E (p. 535), the image I formed by the objective must be situated between F and E. This, then, is the place where cross-wires must be placed if measurement of the final image is required. But the cross-wires are viewed through one lens, E, while the distant object is viewed by rays refracted through both lenses, F, E. The relative lengths of the cross-wires and image are thus rendered disproportionate, and hence cross-wires cannot be used with Huygens' eye-piece, which is a disadvantage.

Ramsden's Eye-piece

Ramsden's eye-piece is more commonly used than Huygens' eye-piece. It consists of two plano-convex lenses of equal focal length f , the distance between them being $2f/3$, Fig. 22.11 (ii). The achromatic condition requires that the distance between the lenses should be f , the average of the focal lengths. If the field lens F were at the focus of the eye lens E, however, E would magnify any dust on F, and vision would then be obscured. F is placed at a distance $f/4$ from the focus of the objective of the telescope, where the real image is formed, in which case the image in F is formed at a distance from E equal to f , its focal length, and parallel rays emerge from E.

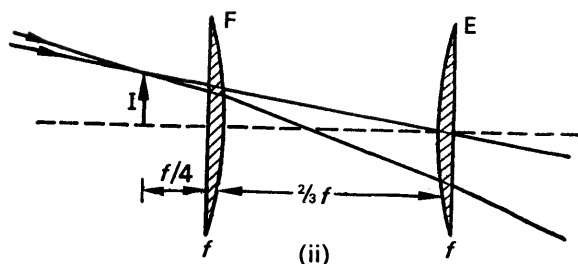


FIG. 22.11 (ii). Ramsden's eye-piece.

The chromatic aberration of Ramsden's eye-piece is small, as is the spherical aberration. The advantage of the eye-piece, however, lies in the fact that cross-wires can be used with it; they are placed outside the combination at the place where the real image I is formed.

EXAMPLES

1. A thin biconvex lens is placed with its principal axis first along a beam of parallel red light and then along a beam of parallel blue light. If the refractive indices of the lens for red and for blue light are respectively 1.514 and 1.524, and if the radii of curvature of its faces are 30 cm and 20 cm, calculate the separation of the foci for red and blue light. What relation does the result bear to the dispersive power of the lens for the two kinds of light? (N .)

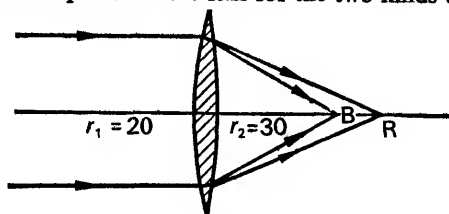


FIG. 22.12.

(a) $\frac{1}{f} = (n - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right)$, with the usual notation.

$$\therefore \frac{1}{f_r} = (1.514 - 1) \left(\frac{1}{+20} + \frac{1}{+30} \right)$$

$$\therefore \frac{1}{f_r} = 0.514 \left(\frac{1}{20} + \frac{1}{30} \right) = \frac{0.514}{12}$$

$$\therefore f_r = 23.33 \text{ cm.}$$

$$\text{Also, } \frac{1}{f_b} = (1.524 - 1) \left(\frac{1}{+20} + \frac{1}{+30} \right) = \frac{0.524}{12}$$

$$\therefore f_b = 22.90 \text{ cm.}$$

$$\therefore \text{separation} = 23.33 - 22.90 = 0.43 \text{ cm.}$$

(b) We know that $\frac{1}{f_b} - \frac{1}{f_r} = \frac{n_b - n_r}{n - 1} \cdot \frac{1}{f}$ (See p. 516.)

$$\text{i.e., } \frac{1}{f_b} - \frac{1}{f_r} = \frac{\omega}{f}$$

$$\therefore \text{dispersive power, } \omega_2, \text{ for glass A} = \frac{n_b - n_r}{n - 1} = \frac{1.5245 - 1.5155}{1.52 - 1} = \frac{9}{520}$$

$$\text{and mean refractive index, } n, \text{ for glass B} = \frac{1.659 + 1.641}{2} = 1.65$$

$$\therefore \text{dispersive power, } \omega_1, \text{ for glass B} = \frac{1.659 - 1.641}{1.65 - 1} = \frac{18}{650}$$

The condition for achromatism is $\frac{f_1}{f_2} = -\frac{\omega_1}{\omega_2}$

$$\therefore \frac{f_1}{f_2} = -\frac{18}{650} \bigg/ \frac{9}{520} \quad \dots \quad (ii)$$

$$\text{From (ii),} \quad f_1 = -\frac{18}{650} \times \frac{520}{9} \cdot f_2 = -\frac{8}{5} f_2$$

Substituting in (i),

$$\therefore -\frac{5}{8f_2} + \frac{1}{f_2} = \frac{1}{100}$$

$$\therefore f_2 = +37.5$$

$$\therefore f_1 = -\frac{8}{5} \times +37.5 = -60$$

Thus a convex lens of 37.5 cm focal length of glass A should be combined with a concave lens of 60 cm focal length of glass B.

EXERCISES 22

Defects of Vision

1. Explain how the eye is focused for viewing objects at different distances. Describe and explain the defects of vision known as *long sight* and *short sight*, and their correction by the use of spectacles.

Explain the advantages we gain by the use of two eyes instead of one.

A certain person can see clearly objects at distances between 20 cm and 200 cm from his eye. What spectacles are required to enable him to see distant objects clearly, and what will be his least distance of distinct vision when he is wearing them? (L.)

2. Give an account of the common optical defects of the human eye and explain how their effects may be corrected.

An elderly person cannot see clearly, without the use of spectacles, objects nearer than 200 cm. What spectacles will he need to reduce this distance to 25 cm? If his eyes can focus rays which are converging to points not less than 150 cm behind them, calculate his range of distinct vision when using the spectacles. (N.)

3. Describe the optical functions of the cornea and lens in the human eye and explain how the corresponding purposes are served in a camera.

In order to correct his near point to 25 cm a man is given spectacles with converging lenses of 50 cm focal length, and to correct his far point to infinity he is given diverging lenses of 200 cm focal length. Ignoring the separation of lens and eye determine the distances of his near and far points when not wearing spectacles and suggest reasons for his defects of vision.

The glass used for the lenses has a refractive index of 1.50 and the back surface of each lens is concave to the eye with a radius of curvature of 50 cm.

Calculate the radii of the front surfaces and state whether the surfaces are convex or concave outwards. (N.)

4. Describe a method of determining the focal length of a thin diverging lens of power numerically about 6 dioptres. Give the theory of the method, assuming any necessary relation between object and image distances, but showing clearly the signs to be attached to the numerical data.

Explain fully (a) the use of diverging lenses to correct a defect of vision, (b) the optical arrangement of a telescope with a diverging lens as eyepiece. (L.)

5. State the sign convention you employ in solving optical problems. Discuss its application to the radii of curvature of mirrors. Write down the formula relating the object distance with the image distance and the radius of curvature for a spherical mirror of small aperture. Using the same convention and notation, write down the formula connecting object and image distance for a plane mirror.

In order to read a book held as close as 20 cm away from his eyes, a man requires spectacles with converging lenses of focal length 22 cm. How far away from him is the closest object he can clearly see without the spectacles? Assume that the distance between the eye and the spectacle lens is negligible. Outline the physics of two possible causes of this defect of vision. (C.)

6. Describe the optical system of the eye and explain the meaning of *long sight*, *short sight*, and *least distance of distinct vision*. Illustrate with clear diagrams the two defects of vision mentioned above and show how they are corrected by lenses.

If the range of vision of a short-sighted man is from 10 to 20 cm from the eye, what lens should be used in order to enable him to see distant objects clearly? What would be the range of accommodation when using this lens? (L.)

7. Describe the optical arrangement of the eye, illustrating the description with a labelled diagram.

A person wears bifocal converging spectacles, one surface of each lens being spherical and the other cylindrical. Describe the defects in his vision and explain how the spectacles correct them. (N.)

8. Give clear diagrams to illustrate the common optical defects of the human eye.

In a certain case the range of distinct vision is found to be limited to objects distant 15 cm to 30 cm from the eye. What lens would be suitable for the distant vision of distant objects, and what would be the nearer limit of distinct vision when this lens is in use? (W.)

Defects of Lenses

9. Define *dispersive power* of a transparent medium.

A crosswire illuminated from behind by a white source of light is placed axially and 30.0 cm from a thin converging lens whose focal length for yellow sodium light is 25.0 cm. If the dispersive power of the glass of the lens for the extreme wavelengths in the light source is 0.0170, find the axial separation of the images formed by light of these wavelengths.

Find also the focal length and the nature of a thin lens, made of glass of dispersive power 0.0270, which in contact with the first lens will form an achromatic doublet. (L.)

10. What is meant by *chromatic aberration*, *dispersion*, *dispersive power*?

Derive the necessary relation between the focal lengths of the component

lenses and the dispersive powers of the glasses of which they are made, for a combination of two thin lenses in contact to be achromatic.

A biconvex lens *A* is combined with a planoconcave lens *B* to form an achromatic pair. The adjacent faces are in contact and have a common radius of 15.34 cm. Find the focal length of the combination, given the following refractive indices:

	<i>Blue</i>	<i>Red</i>	<i>Yellow</i>
Lens <i>A</i>	1.5235	1.5149	1.5192
Lens <i>B</i>	1.6635	1.6463	1.6549 (<i>L.</i>)

11. Explain what is meant by (a) spherical aberration, (b) chromatic aberration. Give an account of ways in which these defects are minimised in optical instruments. (*N.*)

12. What is the condition for two thin lenses to form an achromatic doublet? Draw a diagram of such a combination.

A convex lens of crown glass has a focal length of 40 cm and a dispersive power of 0.025. Find the focal length of a flint glass lens of dispersive power 0.04 which will form an achromatic convex doublet with it. Calculate the focal length of the combination.

13. Explain the dispersion produced by a simple lens, and show how the defect may be corrected.

Why is such correction unnecessary in the case of a simple converging lens used as a magnifying glass held close to the eye? (*O. & C.*)

14. Derive the expression for the focal length of a thin lens in air in terms of the radii of curvature of the faces and the refractive index of the material of the lens.

A white disc, 1 cm in diameter, is placed 100 cm in front of a thin converging lens of 50 cm mean focal length. The refractive index of the material of the lens is 1.524 for red light and 1.534 for violet light. Calculate the diameters of the images for red and violet light and their distance apart. (*W.*)

15. Discuss in general terms the defects in the image formed by a single converging lens in a camera and indicate how they may be remedied.

A camera lens forms an image of the same size as the object when the screen is in a certain position. When the screen is moved 10 cm further from the lens and the object is moved until the image is again in focus, the magnification is found to be 2. What is the focal length of the lens? (*C.*)

16. Describe the colour effects which you would expect to see in the image of a small source of white light formed on a screen by a lens.

Derive the condition that a combination of two thin lenses in contact shall be nearly free from this defect. (*L.*)

17. A thin spherical lens is made from glass with a refractive index n , and its surface have radii of curvature r_1 and r_2 . Find an expression for the focal length f of the lens when it is surrounded by a medium of refractive index unity.

If R_1 , R_2 and F are the corresponding numerical values of r_1 , r_2 and f (i.e. ignoring signs) and $R_1 < R_2$, write down the expressions for F for (a) the two possible types of converging lens, and (b) the two possible types of diverging lens.

A single thin spherical lens is used to form the image of a star on the axis of the lens. State and explain the effect on the image of (i) spherical aberration and (ii) chromatic aberration, and describe how one of these aberrations may be reduced. (*O. & C.*)

chapter twenty-three

Optical instruments

WHEN a telescope or a microscope is used to view an object, the appearance of the final image is determined by the cone of rays entering the eye. A discussion of optical instruments and their behaviour must therefore be preceded by a consideration of the image formed by the eye, and we must now recapitulate some of the points about the eye mentioned in previous pages.

Firstly, the image formed by the eye lens L must appear on the retina R at the back of the eye if the object is to be clearly seen, Fig. 23.1. Secondly, the normal eye can focus an object at infinity (the "far point" of the normal eye), in which case the eye is said to be "unaccommodated". Thirdly, the eye can see an object in greatest detail when it is placed at a certain distance D from the eye, known as the *least distance of distinct vision*, which is about 25 cm, for a normal eye (p. 508). The point at a distance D from the eye is known as its "near point".

Visual Angle

Consider an object O placed some distance from the eye, and suppose θ is the angle in radians subtended by it at the eye, Fig. 23.1. Since

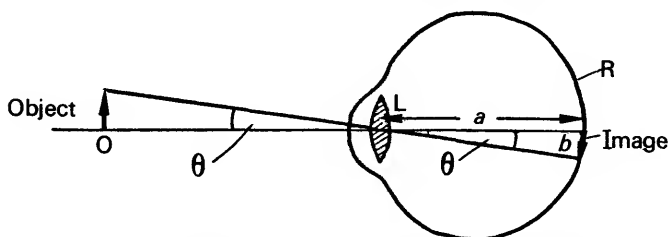


FIG. 23.1. Length of image on retina, and visual angle.

vertically opposite angles are equal, it follows that the length b of the image on the retina is given by $b = a\theta$, where a is the distance from R to L . But a is a constant; hence $b \propto \theta$. We thus arrive at the important conclusion that *the length of the image formed by the eye is proportional to the angle subtended at the eye by the object*. This angle is known as the *visual angle*; the greater the visual angle, the greater is the apparent size of the object.

Fig. 23.2 (i) illustrates the case of an object moved from A to B , and viewed by the eye in both positions. At B the angle β subtended at the eye is greater than the visual angle α subtended at A . Hence the object

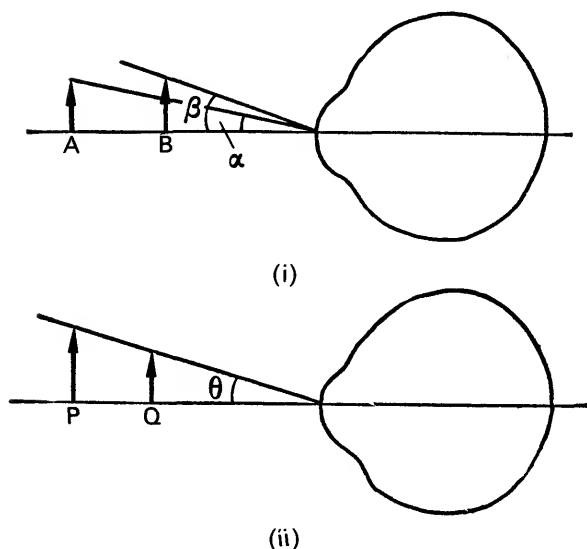


FIG. 23.2. Relation between visual angle and length of image.

appears larger at B than at A, although its physical size is the same. Fig. 23.2 (ii) illustrates the case of two objects, at P, Q respectively, which subtend the same visual angle θ at the eye. The objects thus appear to be of equal size, although the object at P is physically bigger than that at Q. It should be remembered that an object is not clearly seen if it is brought closer to the eye than the near point.

Angular Magnification

Microscopes and telescopes are instruments designed to increase the visual angle, so that the object viewed can be made to appear much larger with their aid. Before they are used the object may subtend a certain angle α at the eye; when they are used the final images may subtend an increased angle α' at the eye. The *angular magnification*, M , of the instrument is defined as the ratio

$$M = \frac{\alpha'}{\alpha} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and this is also popularly known as the *magnifying power* of the instrument. It should be carefully noted that we are concerned with visual angles in the theory of optical instruments, and not with the physical sizes of the object and the image obtained.

Microscopes

At the beginning of the seventeenth century single lenses were developed as powerful magnifying glasses, and many important discoveries in human and animal biology were made with their aid. Shortly afterwards two or more convex lenses were combined to form powerful microscopes, and with their aid Hooke, in 1648, discovered the existence of "cells" in animal and vegetable tissue.

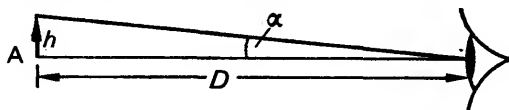
A microscope is an instrument used for viewing *near* objects. When it is in normal use, therefore, the image formed by the microscope is usually at the least distance of distinct vision, D , from the eye, i.e., at the near point of the eye. With the unaided eye (i.e., without the instrument), the object is seen clearest when it is placed at the near point. Consequently the angular magnification of a microscope in *normal* use is given by

$$M = \frac{\alpha'}{\alpha},$$

where α' is the angle subtended at the eye by the image at the near point, and α is the angle subtended at the unaided eye by the object at the near point.

Simple Microscope or Magnifying Glass

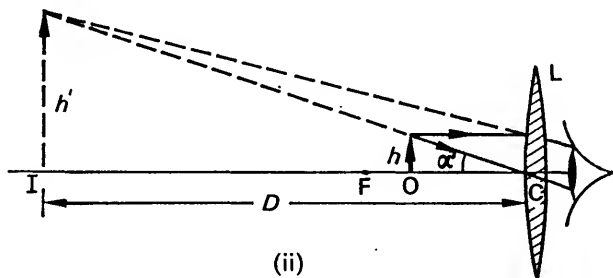
Suppose that an object of length h is viewed at the near point, A, by the unaided eye, Fig. 23.3 (i). The visual angle, α , is then h/D in radian



(i)

FIG. 23.3 (i). Visual angle with unaided eye.

measure. Now suppose that a convex lens L is used as a magnifying glass to view the same object. An erect, magnified image is obtained when the object O is nearer to L than its focal length (p. 486), and the observer moves the lens until the image at I is situated at his near point. If the observer's eye is close to the lens at C, the distance IC is then equal to D , the least distance of distinct vision, Fig. 23.3 (ii). Thus the new visual angle α' is given by h'/D , where h' is the length of the virtual image, and it can be seen that α' is greater than α by comparing Fig. 23.3 (i) with Fig. 23.3 (ii).



(ii)

FIG. 23.3 (ii). Simple microscope, or magnifying glass.

The angular magnification, M , of this simple microscope can be evaluated in terms of D and the focal length f of the lens. From the definition of M (p. 526), $M = \alpha'/\alpha$.

But

$$a' = \frac{h'}{D}, a = \frac{h}{D}$$

$$\therefore M = \frac{h'}{D} \bigg/ \frac{h}{D} = h'/h \quad . \quad . \quad . \quad (i)$$

Now h'/h is the "linear magnification" produced by the lens, and is given by $h'/h = v/u$, where v is the image distance CI and u is the object distance CO (see p. 488). Since $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, with the usual notation, we have

$$1 + \frac{v}{u} = \frac{v}{f},$$

by multiplying throughout by v .

$$\therefore \frac{v}{u} = \frac{v}{f} - 1 = \frac{D}{f} - 1$$

since $v = CI = D$.

$$\therefore \frac{h'}{h} = \frac{D}{f} - 1.$$

$$\therefore M = \frac{D}{f} - 1 \quad . \quad . \quad . \quad (2)$$

from (i) above.

If the magnifying glass has a focal length of 2 cm, $f = +2$ as it is converging; also, if the least distance of distinct vision is 25 cm, $D = -25$ as the image is virtual, see Fig. 23.3 (ii). Substituting in (2),

$$M = \frac{-25}{+2} - 1 = -13\frac{1}{2}.$$

Thus the angular magnification is $13\frac{1}{2}$. The position of the object O is given by substituting $v = -25$ and $f = +2$ in the lens equation

$\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, from which the object distance u is found to be $+1.86$ cm.

From the formula for M in (2), it follows that a lens of *short* focal length is required for high angular magnification.

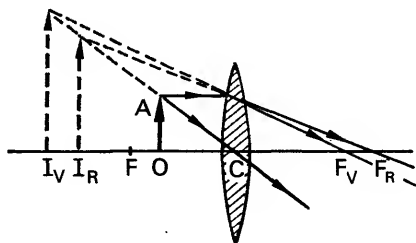


FIG. 23.4. Dispersion with magnifying glass.

When an object OA is viewed through a convex lens acting as a *magnifying glass*, various coloured virtual images, corresponding to I_R, I_V for red and violet rays for example, are formed, Fig. 23.4. The top point of each image lies on the line CA . Hence each image subtends the same angle at the eye close to the lens, so that the colours received by the eye will practically overlap. Thus the virtual image seen in a magnifying glass is almost free of chromatic aberration. A little colour is observed at the edges as a result of spherical aberration. A *real* image formed by a lens has chromatic aberration, as explained on p. 513.

Magnifying Glass with Image at Infinity

We have just considered the normal use of the simple microscope, in which case the image formed is at the near point of the eye and the eye is accommodated (p. 527). When the image is formed at infinity, however, which is not a normal use of the microscope, the eye is undergoing the least strain and is then unaccommodated (p. 508). In this case the object must be placed at the focus, F , of the lens. Fig. 23.5.

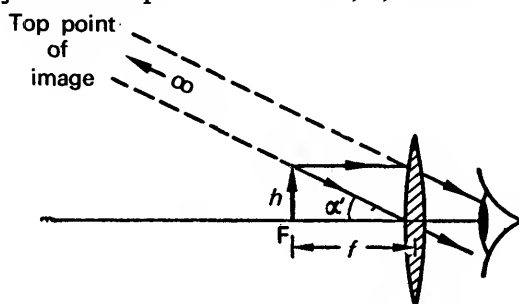


FIG. 23.5. Final image at infinity.

Suppose that the focal length of the lens is f . The visual angle α' now subtended at the eye is then h/f if the eye is close to the lens, and hence the angular magnification, M , is given by

$$M = \frac{\alpha'}{\alpha} = \frac{h/f}{h/D},$$

as $\alpha = h/D$, Fig. 23.3 (i).

$$\therefore M = \frac{D}{f} \quad \dots \dots \dots (3)$$

When $f = +2$ cm and $D = -25$ cm, $M = -12\frac{1}{2}$. The angular magnification was $-13\frac{1}{2}$ when the image was formed at the near point (p. 528). It can easily be verified that the angular magnification varies between $12\frac{1}{2}$ and $13\frac{1}{2}$ when the image is formed between infinity and the near point, and the maximum angular magnification is thus obtained when the image is at the near point.

Compound Microscope

From the formula $M = D/f - 1$, M is greater the smaller the focal length of the lens. As it is impracticable to decrease f beyond a certain

limit, owing to the mechanical difficulties of grinding a lens of short focal length (great curvature), *two* separated lenses are used to obtain a high angular magnification, and constitute a *compound* microscope. The lens nearer to the object is called the *objective*; the lens through which the final image is viewed is called the *eye-piece*. The objective and the eye-piece are both converging, and both have small focal lengths for a reason explained later (p. 531).

When the microscope is used, the object *O* is placed at a slightly *greater* distance from the objective than its focal length. In Fig. 23.6, F_o is the focus of this lens. An inverted real image is then formed at I_1 in the microscope tube, and the eye-piece is adjusted so that a large virtual image is formed by it at I_2 . Thus I_1 is *nearer* to the eye-piece

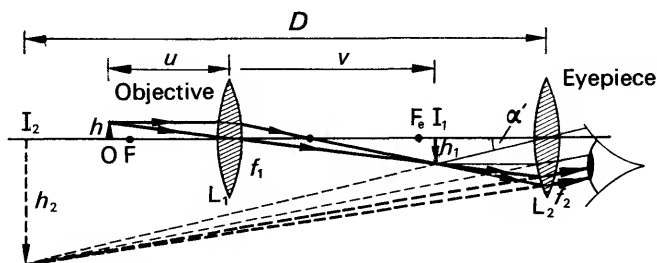


FIG. 23.6. Compound microscope in *normal* use.

than the focus F_e of this lens. It can now be seen that the eye-piece functions as a simple magnifying glass, used for viewing the image formed at I_1 by the objective.

Angular Magnification with Microscope in Normal Use

When the microscope is in normal use the image at I_2 is formed at the least distance of distinct vision, D , from the eye (p. 527). Suppose that the eye is close to the eye-piece, as shown in Fig. 23.6. The visual angle α' subtended by the image at I_2 is then given by $\alpha' = h_2/D$, where h_2 is the height of the image. With the unaided eye, the object subtends a visual angle given by $\alpha = h/D$, where h is the height of the object, see Fig. 23.3 (i).

$$\begin{aligned} \therefore \text{angular magnification, } M &= \frac{\alpha'}{\alpha} \\ &= \frac{h_2/D}{h/D} = \frac{h_2}{h}. \end{aligned}$$

Now $\frac{h_2}{h}$ can be written as $\frac{h_2}{h_1} \times \frac{h_1}{h}$, where h_1 is the length of the intermediate image formed at I_1 .

$$\therefore M = \frac{h_2}{h_1} \cdot \frac{h_1}{h} \quad \dots \quad (i)$$

The ratio h_2/h_1 is the linear magnification of the "object" at I_1 produced by the *eye-piece*, and we have shown on p. 528 that the linear

magnification is also given by $\frac{v}{f_2} - 1$, where v is the image distance from the lens and f_2 is the focal length. Since $v = D$ numerically in this case (the image at I_2 is at a distance $-D$ from the eye-piece), it follows that

$$\frac{h_2}{h_1} = \frac{D}{f_2} - 1 \quad . \quad . \quad . \quad . \quad (\text{ii})$$

Also, the ratio h_1/h is the linear magnification of the object at O produced by the *objective* lens. Thus if the distance of the image I_1 from this lens is denoted by v , we have

$$\frac{h_1}{h} = \frac{v}{f_1} - 1 \quad . \quad . \quad . \quad . \quad (\text{iii})$$

$$\therefore M = \frac{h_2}{h_1} \cdot \frac{h_1}{h} = \left(\frac{D}{f_2} - 1 \right) \left(\frac{v}{f_1} - 1 \right) \quad . \quad . \quad (4)$$

It can be seen that if f_1 and f_2 are small, M is large. Thus the angular magnification is high if the focal lengths of the objective and the eyepiece are small.

Microscope with Image at Infinity

The compound microscope can also be used with the final image formed at infinity, which is not the normal use of the instrument. In this case the eye is unaccommodated, or "at rest". The image of the object in the objective must now be formed at the focus, F_e , of the eye-piece, as shown in Fig. 23.4, and the visual angle α' subtended at the eye by the final image at infinity is then given by $\alpha' = h_1/f_2$, where h_1 is the length of the image at I_1 , and f_2 is the focal length of the eye-piece.

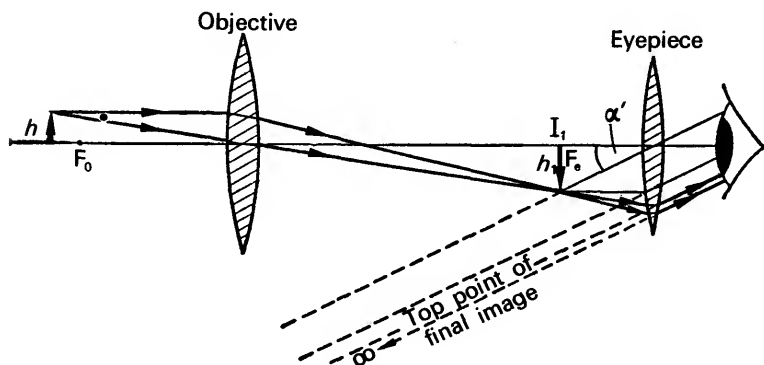


FIG. 23.7. Microscope with image at infinity.

The angular magnification, M , is given by $M = a'/a$, where $a = h/D$ (p. 527).

$$\therefore M = \frac{\alpha'}{\alpha} = \frac{h_1/f_2}{h/D} = \frac{h_1}{h} \cdot \frac{D}{f_2}$$

But, from above, $\frac{h_1}{h} = \frac{v}{f_1} - 1$.

$$\therefore M = \left(\frac{v}{f_1} - 1 \right) \frac{D}{f_2} \quad . \quad . \quad . \quad (5)$$

Comparing equations (4) and (5), it can be seen, since D is a negative (virtual) quantity, that the angular magnification is greater when the final image is formed at the near point. Further, it will be noted that the eye-piece is nearer to the image at I_1 in the latter case.

The Best Position of the Eye. The Eye-Ring

When an object is viewed by an optical instrument, only those rays from the object which are bounded by the perimeter of the objective lens enter the instrument. The lens thus acts as a *stop* to the light from the object. Similarly, the only rays from the image causing the sensation of vision are those which enter the pupil of the eye. The pupil thus acts as a natural stop to the light from the image; and with a given objective, the best position of the eye is one where it collects as much light as possible from that passing through the objective.

Fig. 23.8 illustrates three of the rays from a point X on an object at O placed in front of a compound microscope. Two of the rays are refracted at the boundary of the objective L_1 to pass through Y on the real image at I_1 , while the ray OC_1 through the middle C_1 of the objective passes straight through to Y . The cone of light is then incident on the eye-piece lens L_2 , where it is refracted and forms the point T on the final image, corresponding to X on the object. Now the central ray of the beams of light incident on L_1 from every point on the object passes through C_1 , the centre of the objective lens. The central ray of the emergent beams from the eye-piece L_2 thus passes through the image of C_1 in L_2 . By similar reasoning, we arrive at the conclusion that *all*

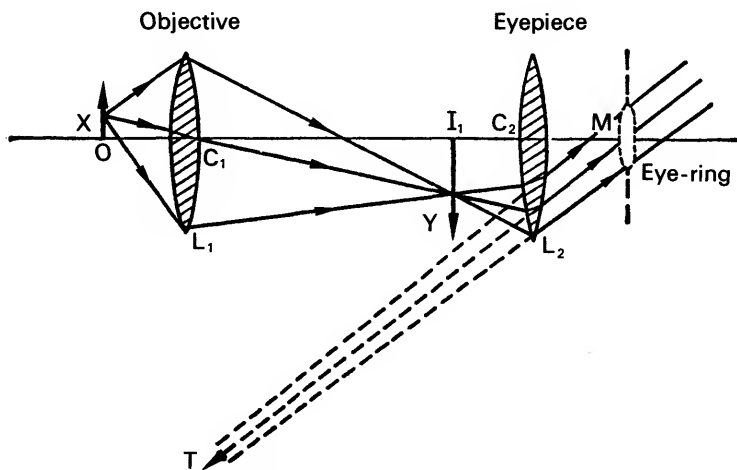


FIG. 23.8. Position of eye-ring.

the emergent rays pass through the image M of the objective in the eye-piece. This image is known as the *eye-ring*, and the best position of the eye is thus at M.

Suppose the objective is 16 cm from L_2 , which has a focal length of 2 cm. The image distance, v , in L_2 is given by $\frac{1}{v} + \frac{1}{(+16)} = \frac{1}{(+2)}$, from which $v = 2.3$ cm. Thus M is a short distance from the eye-piece, and in practice the eye should be farther from the eye-piece than in Fig. 23.8. This is arranged in commercial microscopes by having a circular opening fixed at the eye-ring distance from the eye-piece, so that the observer's eye has automatically the best position when it is placed close to the opening.

Angular Magnification of Telescopes

Telescopes are instruments used for viewing distant objects, and they are used extensively at sea and at astronomical observatories. The first telescope is reputed to have been made about 1608, and in 1609 Galileo made a telescope through which he observed the satellites of Jupiter and the rings of Saturn. The telescope thus paved the way for great astronomical discoveries, particularly in the hands of KEPLER. Newton also designed telescopes, and was the first person to suggest the use of curved mirrors for telescopes free from chromatic aberration (see p. 513).

If α is the angle subtended at the unaided eye by a distant object, and α' is the angle subtended at the eye by its image when a telescope is used, the angular magnification M of the instrument is given by

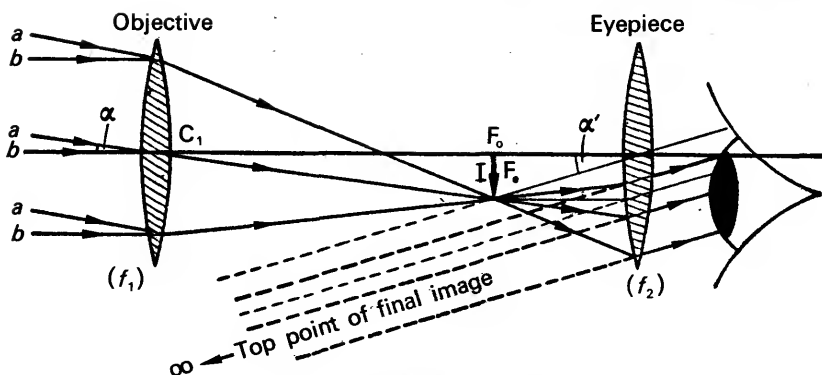
$$M = \frac{\alpha'}{\alpha}.$$

It should be carefully noted that α is *not* the angle subtended at the unaided eye by the object at the near point, as was the case with the microscope, because the telescope is used for viewing distant objects.

Astronomical Telescope in Normal Adjustment

An astronomical telescope made from lenses consists of an objective of long focal length and an eye-piece of short focal length, for a reason given on p. 534. Both lenses are converging. *The telescope is in normal adjustment when the final image is formed at infinity*, and the eye is then unaccommodated when viewing the image. The unaided eye is also unaccommodated when the distant object is viewed, as the latter may be considered to be at infinity.

Fig. 23.9 illustrates the formation of the final image when the telescope is used normally. The image I of the distant object is formed at the focus, F_0 , of the objective since the rays incident on the latter are parallel; and since the final image is formed at infinity, the focus F_e of the eye-piece must also be at F_0 . Fig. 23.9 shows three of the many rays from the *top* point of the object, marked a , and three of the many rays from the *foot* of the object, marked b . These rays pass respectively through the top and foot of the image I, as shown.

FIG. 23.9. Telescope in *normal* use.

We can now obtain an expression for the angular magnification, M , of the telescope; in so doing we shall assume that the eye is close to the eye-piece. Since the length between the objective and the eye-piece is very small compared with the distance of the object from either lens, we can take the angle α subtended at the unaided eye by the object as that subtended at the objective lens, Fig. 23.9. The angle α' subtended at the eye when the telescope is used is given by $\alpha' = h/f_2$, where h is the length of the image I and f_2 is the focal length of the eye-piece.

But

$$\alpha = h/f_1,$$

where f_1 is the focal length of the objective, since I is at a distance f_1 from C_1 .

$$\therefore M = \frac{\alpha'}{\alpha} = \frac{h/f_2}{h/f_1}$$

$$\therefore M = \frac{f_1}{f_2} \quad \dots \quad (6)$$

Thus the angular magnification is equal to the ratio of the focal length of the objective (f_1) to that of the eye-piece (f_2). For high angular magnification, it follows from (6) that the objective should have a long focal length and the eye-piece a short focal length.

It will be noted that the distance between the lenses is equal to the sum ($f_1 + f_2$) of their focal lengths. This provides a simple method of setting up two convex lenses to form an astronomical telescope when their focal lengths are known.

The Eye-ring, and Relation to Angular Magnification

As we explained in the case of the microscope, the rays which pass through the telescope from the distant object are those bounded by the objective lens. Fig. 23.10 illustrates three rays from a point on the distant object which pass through the objective, forming an image at Y . The eye-ring, M , the best position for the eye, is the circular image of the objective in the eye-piece L_2 , and we can calculate its position as

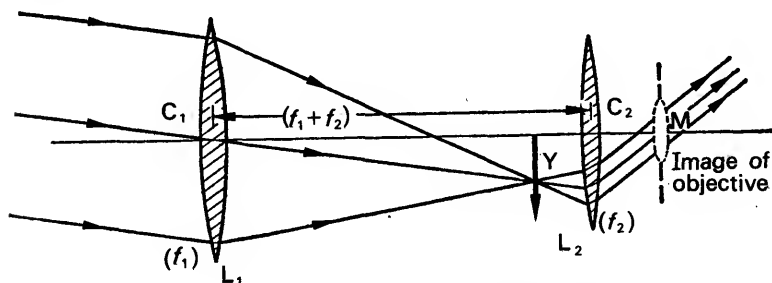


FIG. 23.10. Eye-ring relation to magnification.

$C_1C_2 = f_1 + f_2$, from previous. As the focal length of $L_2 = f_2$, the distance C_2M , or v , is given by $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$,

$$\text{i.e., } \frac{1}{v} + \frac{1}{f_1 + f_2} = \frac{1}{f_2}$$

from which
$$v = \frac{f_2}{f_1} (f_1 + f_2).$$

Now the objective diameter: eye-ring diameter = $C_1C_2 : C_2M$

$$= u : v = (f_1 + f_2) : \frac{f_2}{f_1} (f_1 + f_2) \\ = f_1 / f_2.$$

But the angular magnification of the telescope = f_1/f_2 (p. 534). Thus the angular magnification, M , is also given by

$$M = \frac{\text{diameter of objective}}{\text{diameter of eye-ring}} \quad (7)$$

the telescope being in normal adjustment.

Telescope with Final Image Near Point

When a telescope is used, the final image can be formed at the near point of the eye instead of at infinity. The eye is then "accommodated", and although the image is still clearly seen, the telescope is *not* in normal adjustment (p. 533). Fig. 23.11 illustrates the formation of the final image.

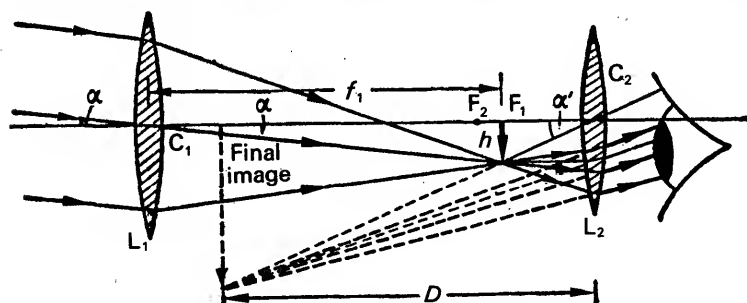


FIG. 23.11. Final image at near point.

The objective forms an image of the distant object at its focus F_1 , and the eye-piece is moved so that the image is nearer to it than its focus F_2 , thus acting as a magnifying glass.

The angle α subtended at the unaided eye is practically that subtended at the objective L_1 . Thus $\alpha = h/f_1$, where h is the length of the image in the objective and f_1 its focal length. The angle α' subtended at the eye by the final image $= h/u$, if the eye is close to the eye-piece, where $u = F_1C_2$ = the distance of the image at F_1 from the eye-piece.

Thus angular magnification, $M = \frac{\alpha'}{\alpha} = \frac{h/u}{h/f_1}$

$$\therefore M = \frac{f_1}{u} \quad \dots \quad (i)$$

As the final image is formed at a numerical distance D from the eye-piece L_2 , we have $v = -D$ when $f = +f_2$. Thus, from $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$,

$$\frac{1}{-D} + \frac{1}{u} = \frac{1}{+f_2},$$

from which

$$u = \frac{f_2 D}{f_2 + D}.$$

Substituting in (i) for u ,

$$\therefore M = \frac{f_1}{f_2} \left(\frac{f_2 + D}{D} \right)$$

$$\therefore M = \frac{f_1}{f_2} \left(1 + \frac{f_2}{D} \right) \quad \dots \quad (8)$$

The angular magnification when the telescope is in normal adjustment (i.e., final image at infinity) $= \frac{f_1}{f_2}$ (p. 534). Hence, from (8), the

angular magnification is increased in the ratio $\left(1 + \frac{f_2}{D} \right) : 1$ when the final image is formed at the near point.

Terrestrial Telescope

From Fig. 23.11, it can be seen that the top point of the distant object is above the axis of the lens, but the top point of the final image is below the axis. Thus the image in an astronomical telescope is *inverted*. This instrument is suitable for astronomy because it makes little difference if a star, for example, is inverted, but it is useless for viewing objects on the earth or sea, in which case an erect image is required.

A *terrestrial telescope* provides an erect image. In addition to the objective and eye-piece of the astronomical telescope, it has a converging lens L of focal length f between them, Fig. 23.12. L is placed at a distance $2f$ in front of the inverted real image I_1 formed by the objective, in which case, as shown on p. 488, the image I in L of I_1 (i) is inverted, real, and the same size as I_1 , (ii) is also at a distance $2f$ from L . Thus the image I is now the same way up as the distant object. If I is

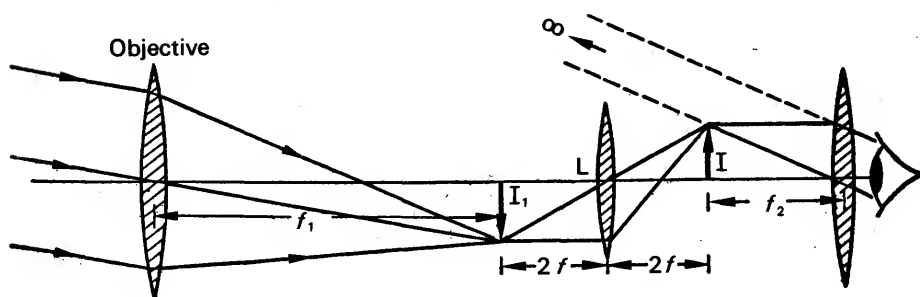


FIG. 23.12. Terrestrial telescope.

at the focus of the eye-piece, the final image is formed at infinity and is also erect.

The lens L is often known as the “erecting” lens of the telescope, as its only function is that of inverting the image I_1 . Since the image I produced by L is the same size as I_1 , the presence of L does not affect the magnitude of the angular magnification of the telescope, which is thus f_1/f_2 (p. 538). The erecting lens, however, reduces the intensity of the light emerging through the eye-piece, as light is reflected at the lens surfaces. Yet another disadvantage is the increased length of the telescope when L is used; the distance from the objective to the eye-piece is now $(f_1 + f_2 + 4f)$, Fig. 23.12, compared with $(f_1 + f_2)$ in the astronomical telescope.

Galileo's Telescope

About 1610; with characteristic genius, Galileo designed a telescope which provides an erect image of an object with the aid of only two lenses. The *Galilean telescope* consists of an objective which is a converging lens of long focal length, and an eye-piece which is a *diverging* lens of short focal length. The distance between the lenses is equal to the *difference* in the magnitudes of their focal lengths, i.e., $C_1C_2 =$

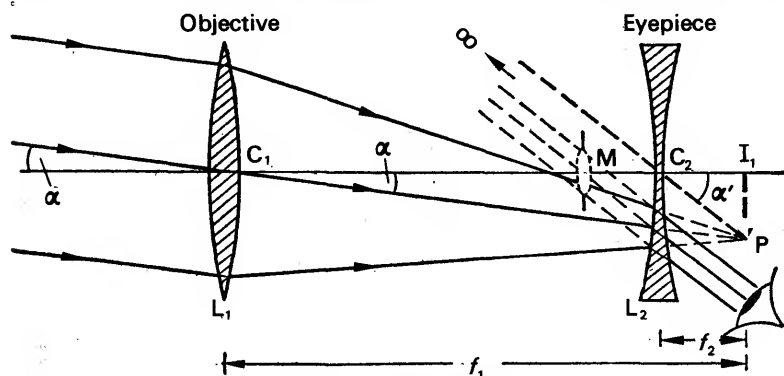


FIG. 23.13. Galilean telescope.

since $C_2I_2 = D$, the least distance of distinct vision, we have $v = -D$ (the image in L_2 is virtual) and f_2 is negative. Since $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$, we obtain

$$\frac{1}{-D} + \frac{1}{u} = \frac{1}{-f_2},$$

assuming f_2 is the numerical value of the diverging lens focal length, from which $u = \frac{-f_2 D}{D - f_2}$.

With the usual notation, the angular magnification, $M = \frac{\alpha'}{\alpha}$.

But $\alpha' = h/u$, $\alpha = h/f_1$. Thus $M = f_1/u$.

But

$$u = \frac{-f_2 D}{D - f_2}$$

$$\therefore M = \frac{f_1}{f_2} \left(\frac{f_2}{D} - 1 \right)$$

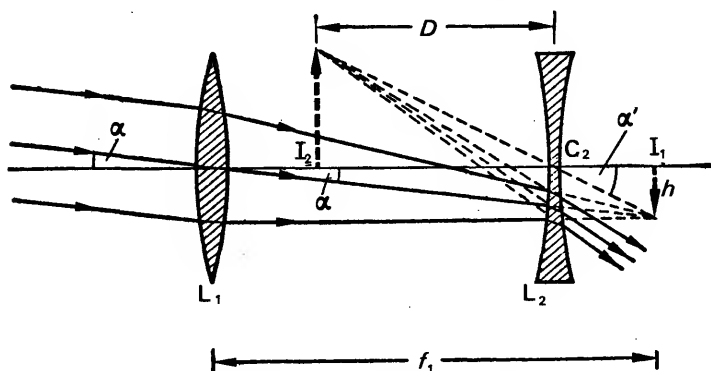


FIG. 23.14. Final image at near point.

Measurement of magnifying power of telescope and microscope

Method 1. The magnifying power of a telescope can be measured by placing a well-illuminated large scale S at one end of the laboratory, and viewing it through the telescope at the other end. If the telescope consists of converging lenses O , E acting as objective and eye-piece respectively, the distance

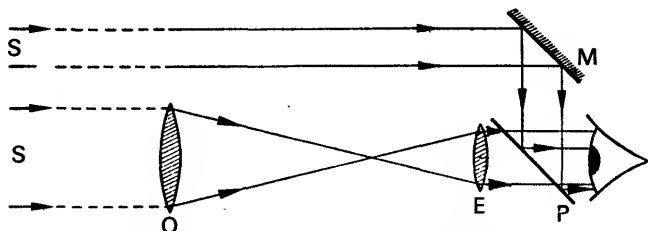


FIG. 23.15. Measurement of magnifying power.

between O, E is the sum of their focal lengths when the instrument is in normal use (p. 534), Fig. 23.15. By means of a plane mirror M and a plane piece of glass P, the divisions on the scale S can be superimposed on its image seen through the telescope, and the ratio of the divisions in equal lengths of the field of view can thus be determined. Since the ratio of the angles subtended at the eye by the image and by the scale is equal to the ratio, a , of the divisions, the magnifying power is equal to a . There must be no parallax between the scale and the final image.

Method 2. The magnifying power of a telescope is the ratio of the diameters of the objective and eye-ring (p. 535). The eye-ring is the image of the objective in the eye-piece, and is obtained by pointing the telescope to the sky and holding a ground glass screen near the eye-piece. A circle, which is the image of the objective in the eye-piece, is observed on the screen, and its diameter, d , is measured. The magnifying power is then given by the ratio d_0/d , where d_0 is the diameter of the objective.

This method is particularly useful when the telescope is fixed in a tube, as it is in practice. When a telescope is set up as in Fig. 23.10, a piece of cardboard with a circular hole can be placed round the objective lens to define its diameter, and the eye-ring found by placing a screen near the eye-piece.

The magnifying power of a microscope can be found by placing two similar scales in front of it, one being 25 cm from the eye. The other scale is placed near the objective, and the eye-piece is moved until the image of this scale coincides with the first scale by the method of no parallax, both eyes being used. The magnifying power is then given by the ratio of the number of divisions occupying the same length.

Prism Binoculars

Prism binoculars are widely used as field glasses, and consist of short astronomical telescopes containing two right-angled isosceles prisms between the objective and eye-piece, Fig. 23.16. These lenses are both converging, and they would produce an inverted image of the distant object if they were alone used. The purpose of the two prisms is to invert the image and obtain a final *erect* image.

One prism A, is placed with its refracting edge vertical, while the other, B, is placed with its refracting edge horizontal. As shown in

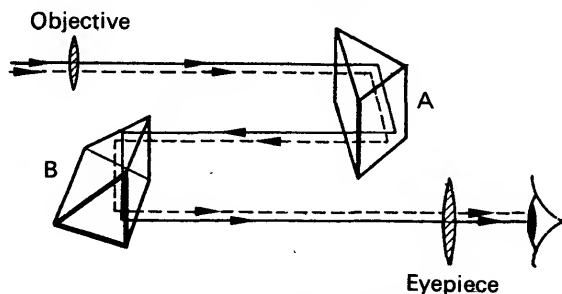


FIG. 23.16. Prism binoculars.

Fig. 23.11, the image formed by the objective alone is inverted. Prism A, however, turns it round in a horizontal direction, and prism B inverts

it in a vertical direction, both prisms acting as reflectors of light (see p. 449). The image produced after reflection at B is now the same way up, and the same way round, as the original object. Since the eye-piece is a convex lens acting as a magnifying glass, it produces a final image the same way up as the image in front of it, and hence the final image is the same way up as the distant object.

Fig. 23.16 illustrates the path of two rays through the optical system. Since the optical path of a ray is about 3 times the distance d between the objective and the eye-piece, the system is equivalent optically to an astronomical telescope of length $3d$. The focal lengths of the objective and eye-piece in the prism binoculars can thus provide the same angular magnification as an astronomical telescope 3 times as long. The compactness of the prism binocular is one of its advantages; another advantage is the wide field of view obtained, as it is an astronomical telescope (p. 541).

Projection Lantern

The projection lantern is used for showing slides on a screen, and the essential features of the apparatus are illustrated in Fig. 23.17. S is a slide whose image is formed on the screen A by adjusting the position of an achromatic objective lens L.

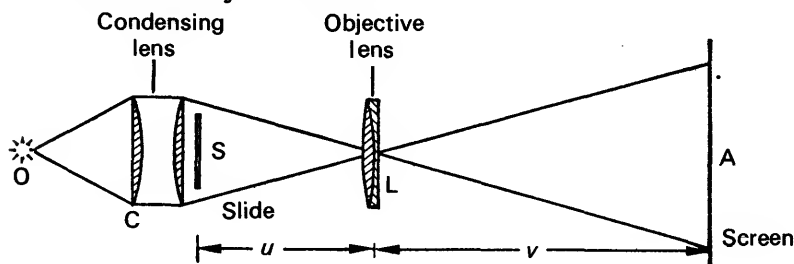


FIG. 23.17. Projection lantern.

The illumination of the slide must be as high as possible, otherwise the image of it on the screen is difficult to see clearly. For this purpose a very bright point source of light, O, is placed near a *condensing lens* C, and the slide S is placed immediately in front of C. The condensing lens consists of a plano-convex lens arrangement, which concentrates the light energy from O in the direction of S, and it has a short focal length. The lens L and the source O are arranged to be conjugate foci for the lens C (i.e., the image of O is formed at L), in which case (i) all the light passes through L, and (ii) an image of O is not formed on the screen. Fig. 23.17 illustrates the path of the beam of light from O which forms the image of S on the screen.

The linear magnification, m , of the slide is given by $m = \frac{v}{u}$, where v, u are the respective screen and slide distances from L. Now $\frac{v}{u} = \frac{v}{f} - 1$ (see p. 528). Thus the required high magnification is obtained by using an objective whose focal length is small compared with v .

Pinhole Camera

The *pinhole camera* consists essentially of a closed box with a pinhole in front and a photographic plate at the back on which the image is formed. The principle was first discovered by PORTA about 1600, who found that clear images were formed on a screen at the back of the box when objects were placed in front of the pinhole.

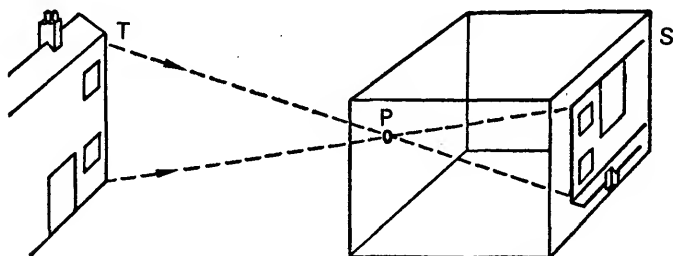


FIG. 23.18. Action of pinhole camera (*not to scale*).

The simple camera utilises the principle that rays of light normally travel in straight lines. As the pinhole P is small, a very narrow cone of rays pass through P from a point T on a house, for example, in front of the box; thus a well-defined image of T is obtained on the photographic plate S , Fig. 23.18. Similarly, other points on the building give rise to clear images on the screen. If the pinhole is enlarged a blurred image is obtained, as the rays from different points on the building then tend to overlap.

The pinhole camera is used by surveyors to photograph the outline of buildings, as the image obtained is free from the distortion produced by the lens in a normal camera.

Photographic Camera; f -number

The photographic camera consists essentially of a *lens system* L , a *light-sensitive film* F at the back, and a *focusing arrangement*, Fig. 23.19.

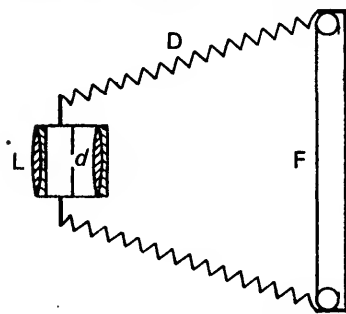


FIG. 23.19. Photographic camera.

The latter is usually a concertina-shaped canvas bag D , which adjusts the distance of the lens L from F . The lens in the camera is an achromatic doublet (p. 515), and the use of two lenses diminishes spherical aberration (p. 518). An *aperture* or *stop* of diameter d is provided, so that the light is incident centrally on the lens, thus diminishing distortion.

The amount of luminous flux falling on the image in a camera is proportional to the area of the lens aperture, or to d^2 , where d is the diameter of the aperture. The area of the image formed is proportional to f^2 , where f is the focal length of

the lens, since the length of the image formed is proportional to the focal length, as illustrated by Fig. 23.20. It therefore follows that the

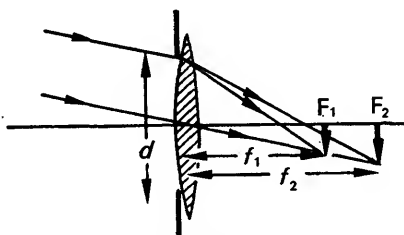


FIG. 23.20. Brightness of image.

luminous flux per unit area of the image, or *brightness* B , of the image, is proportional to d^2/f^2 . The time of exposure, t , for activating the chemicals on the given negative is inversely proportional to B . Hence

$$t \propto \frac{f^2}{d^2} \quad \dots \quad (i)$$

The *relative aperture* of a lens is defined as the ratio d/f , where d is the diameter of the aperture and f is the focal length of the lens. The aperture is usually expressed by its *f-number*. If the aperture is $f/4$, this means that the diameter d of the aperture is $f/4$, where f is the focal length of the lens. An aperture of $f/8$ means a diameter d equal to $f/8$, which is a smaller aperture than $f/4$.

Since the time t of exposure is proportional to f^2/d^2 , from (i), it follows that the exposure required with an aperture $f/8$ ($d = f/8$) is 16 times that required with an aperture $f/2$ ($d = f/2$). The f -numbers on a camera are 2, 2.8, 3.5, 4, 4.8, for example. On squaring the values of f/d for each number, we obtain the relative exposure times, which are 4, 8, 12, 16, 20, or 1, 2, 3, 4, 5.

Depth of Field

An object will not be seen by the eye until its image on the retina covers at least the area of a single cone, which transmits along the optic nerve light energy just sufficient to produce the sensation of vision. As a basis of calculation in photography, a circle of finite diameter about 0.25 mm viewed 250 mm away will just be seen by the eye, as a fairly sharp point, and this is known as the *circle of least confusion*. It corresponds to an angle of about $1/1000$ th radian subtended by an object at the eye.

On account of the lack of resolution of the eye, a camera can take clear pictures of objects at different distances. Consider a point object O in front of a camera lens A which produces a point image I on a film, Fig. 23.21. If XY represents the diameter of the circle of least confusion round I , the eye will see all points in the circle as reasonably sharp points. Now rays from the lens aperture to the edge of XY meet at I_1 beyond I , and also at I_2 in front of I . The point images I_1, I_2 correspond

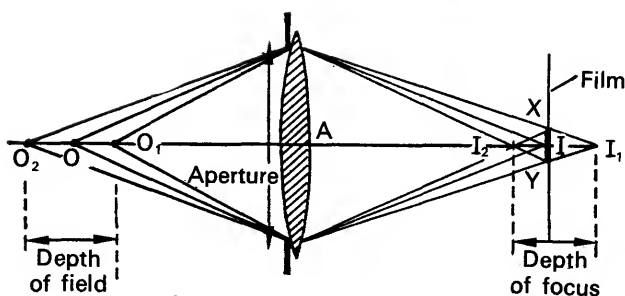


FIG. 23.21. Depth of field.

to point objects O_1 , O_2 on either side of O , as shown. Consequently the images of all objects between O_1 , O_2 are seen clearly on the film.

The distance O_1O_2 is therefore known as the *depth of field*. The distance I_1I_2 is known as the *depth of focus*. The depth of field depends on the lens aperture. If the aperture is made smaller, and the diameter XY of the circle of least confusion is unaltered, it can be seen from Fig. 23.21 that the depth of field increases. If the aperture is made larger, the depth of field decreases.

The Mount Palomar Telescope

The construction of the largest telescope in the world is one of the most fascinating stories of scientific skill and invention. The major feature of the telescope is a parabolic *mirror*, 5 metres across, which is made of pyrex, a low expansion glass. The glass itself took more than six years to grind and polish, having been begun in 1936, and the front of the mirror is coated with aluminium, instead of being covered with silver, as it lasts much longer. The huge size of the mirror enables enough light from very distant stars and planets to be collected and brought to a focus for them to be photographed. Special cameras are incorporated in the instrument to photograph the universe. This method has the advantage that plates can be exposed for hours, if necessary, to the object to be studied, enabling records to be made. It is used to obtain useful information about the building-up and breaking-down of the elements in space (thus assisting in atomic energy research), to investigate astronomical theories of the universe, and to photograph Mars.

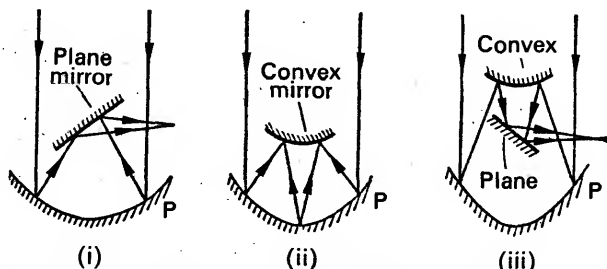


FIG. 23.22. (i). Newton reflector. (ii). Cassegrain reflector. (iii). Coudé reflector.

Besides the main parabolic mirror P, seven other mirrors are used in the 5 metre telescope. Some are plane, Fig. 23.22 (i), while others are convex, Fig. 23.22 (ii), and they are used to bring the light to a more convenient focus, where the image can be photographed, or magnified several hundred times by an eye-piece for observation. The various methods of focusing the image were suggested respectively by *Newton*, *Cassegrain*, and *Coudé*, the last being a combination of the former two methods.

EXAMPLES

1. What do you understand by the magnifying power of an astronomical telescope? Illustrate your answer with a ray diagram depicting the use of the instrument to view stars in the heavens. If such a telescope has an object glass of focal length 50 cm and an eye lens of focal length 5 cm, what is its magnifying power? If it is assumed that the eye is placed very close to the eye lens and that the pupil of the eye has a diameter of 3 mm, what will be the diameter of the object glass if all the light passing through the object glass is to emerge as a beam which fills the pupil of the eye? Assume that the telescope is pointed directly at a particular star. (*W.*)

First part. See text.

Second part. Assuming the telescope is in *normal* adjustment, the final image is formed at infinity. The magnifying power of the telescope is then $\frac{50 \text{ cm}}{5 \text{ cm}}$, or 10. See p. 529.

If all the light emerging from the eye-piece fills the pupil of the eye, the pupil is at the eye-ring. See p. 534. The eye-ring is the image of the objective in the eye-piece. Since the distance, u , from the objective to the eye-piece = 50 + 5 = 55 cm, the eye-ring distance, v , is given by

$$\frac{1}{v} + \frac{1}{(+55)} = \frac{1}{(+5)}$$

from which

$$v = 5.5 \text{ cm.}$$

This is the position of the pupil of the eye. The magnification of the objective is given by

$$\frac{\text{eye-ring diameter}}{\text{objective diameter}} = \frac{v}{u} = \frac{5.5}{55} = \frac{1}{10}$$

$$\therefore \frac{3 \text{ mm}}{\text{objective diameter}} = \frac{1}{10}$$

$$\therefore \text{objective diameter} = 3 \text{ cm.}$$

2. What do you understand by (a) the apparent size of an object, and (b) the magnifying power of a microscope? A model of a compound microscope is made up of two converging lenses of 3 and 9 cm focal length at a fixed separation of 24 cm. Where must the object be placed so that the final image may be at infinity? What will be the magnifying power if the microscope as thus arranged is used by a person whose nearest distance of distinct vision is 25 cm? State what is the best position for the observer's eye and explain why. (*L.*)

First part. (a) The apparent size of an object is proportional to the visual angle. See p. 525. (b) The magnifying power of a microscope is defined on p. 527.

Second part. (i) Suppose the objective A is 3 cm focal length, and the eye-piece B is 9 cm focal length, Fig. 23.23. If the final image is at infinity, the image I_1 in the objective must be 9 cm from B, the focal length of the eye-piece. See p. 531. Thus the image distance LI_1 , from the objective A = $24 - 9 = 15$ cm. The object distance OL is thus given by

$$\frac{1}{(+15)} + \frac{1}{u} = \frac{1}{(+3)},$$

from which

$$u = OL = 3\frac{3}{4} \text{ cm.}$$

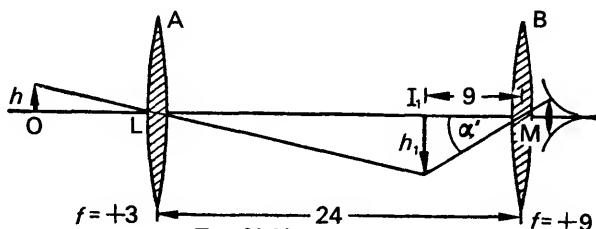


FIG. 23.23. Example.

(ii) The angle α' subtended at the observer's eye is given by $\alpha' = h_1/9$, where h_1 is the height of the image at I_1 , Fig. 23.23. Without the lenses, the object subtends an angle α at the eye given by $\alpha = h/25$, where h is the height of the object, since the least distance of distinct vision is 25 cm.

$$\therefore \text{magnifying power } M = \frac{\alpha'}{\alpha} = \frac{h_1/9}{h/25} = \frac{25}{9} \times \frac{h_1}{h}$$

But

$$\frac{h_1}{h} = \frac{LI_1}{LO} = \frac{15}{3\frac{3}{4}} = 4$$

$$\therefore M = \frac{25}{9} \times 4 = 11\frac{1}{9}$$

The best position of the eye is at the eye-ring, which is the image of the objective A in the eye-piece B (p. 534).

3. A Galilean telescope has an object-glass of 12 cm focal length and an eye lens of 5 cm focal length. It is focused on a distant object so that the final image seen by the eye appears to be situated at a distance of 30 cm from the eye lens. Determine the angular magnification obtained and draw a ray diagram. What are the advantages of prism binoculars as compared with field glasses of the Galilean type? (N.)

Suppose I_2 is the final image, distant 30 cm from the eye lens L_2 . Fig. 23.24. The corresponding object is I_1 . Since I_2 is a virtual image in L_2 , $v = I_2L_2 = -30$, $f = -5$, $u = L_2I_1$. From the lens equation for L_2 , we have

$$\frac{1}{(-30)} + \frac{1}{u} = \frac{1}{(-5)}$$

from which

$$u = -6 \text{ cm.}$$

Thus I_1 is a virtual object for L_2 .

The angular magnification, M , is given by $M = \alpha'/\alpha$. Now $\alpha' = h_1/L_2I_1$, and $\alpha = h_1/L_1I_1$.

$$\therefore M = \frac{h_1/L_2I_1}{h_1/L_1I_1} = \frac{L_1I_1}{L_2I_1}$$

But $L_2I_1 = 6$ cm, from above, and $L_1I_1 =$ focal length of $L_1 = 12$ cm, since the object is distant.

$$\therefore M = \frac{12}{6} = 2$$

The advantages of the prism binoculars are given on p. 541.

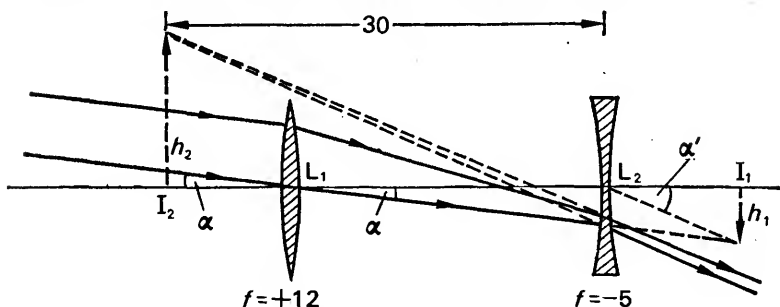


FIG. 23.24. Example.

4. Describe the optical system of a projection lantern. A lantern is required for the projection of slides 7.5 cm square on to a screen 2.1 m square. The distance between the front of the lantern and the screen is to be 20 m. What focal length of projection lens would you consider most suitable?

First part. See text.

Second part. Suppose O is the slide, L is the projection lens, and S is the screen, Fig. 23.25. The linear magnification, m , due to the lens is given by

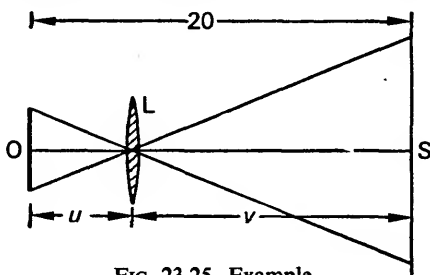


FIG. 23.25. Example.

$$m = \frac{210 \text{ cm}}{7.5 \text{ cm}} = 28$$

$$\therefore LS : LO = 28 : 1$$

$$\therefore LS = v = \frac{28}{29} \times 20 \text{ m}$$

$$\text{and } LO = u = \frac{1}{29} \times 20 \text{ m.}$$

Applying the lens equation,

$$\therefore \frac{1}{560/29} + \frac{1}{20/29} = \frac{1}{f}$$

from which

$$f = \frac{560}{841} \text{ m} = 67 \text{ cm (to nearest cm)}$$

EXERCISES 23

1. An object is viewed with a normal eye at (i) the least distance of distinct vision, (ii) 40 cm, (iii) 100 cm from the eye. Find the ratio of the visual angles in the three cases, and draw diagrams in illustration.

2. Where should the final image be formed when (a) a telescope, (b) a microscope is in *normal* use? Define the angular magnification (magnifying power) of a telescope and a microscope.

3. Explain the essential features of the astronomical telescope. Define and deduce an expression for the magnifying power of this instrument.

A telescope is made of an object glass of focal length 20 cm and an eyepiece of 5 cm, both converging lenses. Find the magnifying power in accordance with your definition in the following cases: (a) when the eye is focused to receive parallel rays, and (b) when the eye sees the image situated at the nearest distance of distinct vision which may be taken as 25 cm. (L.)

4. Explain the action of a microscope consisting of two thin lenses and show on a diagram the paths of three rays from a non-axial object point to the eye. Distinguish clearly between rays and construction lines.

A microscope having as eyepiece a thin lens of focal length 5.0 cm is set up by an observer whose least distance of distinct vision is 25.0 cm. An observer with defective eyesight has to withdraw the eyepiece by 0.50 cm in order that the image may be at his least distance of distinct vision. Find the nature of the defect and specify the nature and the focal length of the spectacle lens he needs to make his least distance of distinct vision 25.0 cm. (L.)

5. Describe, with the help of diagrams, how (a) a single biconvex lens can be used as a magnifying glass, (b) two biconvex lenses can be arranged to form a microscope. State (i) one advantage, (ii) one disadvantage, of setting the microscope so that the final image is at infinity rather than at the near point of the eye.

A centimetre scale is set up 5 cm in front of a biconvex lens whose focal length is 4 cm. A second biconvex lens is placed behind the first, on the same axis, at such a distance that the final image formed by the system coincides with the scale itself and that 1 mm in the image covers 2.4 cm in the scale. Calculate the position and focal length of the second lens. (O. & C.)

6. Compare and contrast the optical properties of (a) an astronomical telescope, (b) a Galilean telescope, (c) a reflecting telescope. Draw ray diagrams for (a) and (b) showing the path through each instrument of a non-axial pencil of rays from a distant object.

Describe one method by which the image in an astronomical telescope may be made erect. (L.)

7. What is the *eye-ring* of a telescope?

For an astronomical telescope in normal adjustment deduce expressions for the size and position of the eye-ring in terms of the diameter of the object glass and the focal lengths of the object glass and eye-lens.

Discuss the importance of (i) the magnitude of the diameter of the object glass, (ii) the structure of the object glass, (iii) the position of the eye. (L.)

8. Show, by means of a ray diagram, how an image of a distant extended object is formed by an astronomical refracting telescope in normal adjustment (i.e. with the final image at infinity).

A telescope objective has focal length 96 cm and diameter 12 cm. Calculate the focal length and minimum diameter of a simple eyepiece lens for use with the telescope, if the magnifying power required is $\times 24$, and all the light transmitted by the objective from a distant point on the telescope axis is to fall on the eyepiece. Derive any formulae you use.

If the eyepiece is an equiconvex lens made from glass of refractive index 1.6, calculate the radius of curvature of its faces and the minimum thickness of the lens at its centre. (O. & C.)

9. Draw the path of two rays, from a point on an object, passing through the optical system of a compound microscope to the final image as seen by the eye.

If the final image formed coincides with the object, and is at the least distance of distinct vision (25 cm) when the object is 4 cm from the objective, calculate the focal lengths of the objective and eye lenses, assuming that the magnifying power of the microscope is 14. (L.)

10. Define *magnifying power* (*angular magnification*) of an optical instrument.

An astronomical telescope consists of two thin converging lenses which are 25.00 cm apart when the telescope is in normal adjustment. The distance between the lenses is reduced to 24.50 cm and a virtual image of an infinitely distant object is then formed 28.00 cm from the eye lens. Calculate the values of the focal lengths of the two lenses and the magnifying power of the instrument with this adjustment, supposing the eye to be placed at the eye lens.

For this same adjustment show on a labelled diagram (not to scale) the relative positions of the principal foci of the two lenses and the construction lines showing the relation of the final image to the intermediate image. On the separate diagram show the paths through the instrument of two rays from a non-axial point. One of the rays should pass through the centre of the objective and the other through its periphery. (L.)

11. Distinguish between the *magnification* produced by, and the *magnifying power* of, an optical system.

Draw a ray diagram showing the action of a simple astronomical telescope (assume two lenses only) in forming separate images of two stars which are close together and near the axis, the final images being at infinity.

If the objective has a diameter of 30 cm and a focal length of 3 metres, and the focal length of the eyepiece is 1.2 cm, calculate (a) the magnifying power of the telescope, (b) the diameter of the image of the objective formed by the eyepiece. (O. & C.)

12. Briefly describe an optical instrument which includes a reflecting prism. What is the function of the prism and what is the principle governing its action?

What are the advantages in using a prism rather than a silvered mirror in the apparatus you describe?

Parallel rays of light fall normally on the face AC of a total reflection prism ABC of refractive index 1.5 which has angle A exactly 45° , angle B approximately 90° and angle C approximately 45° . After total internal reflections in the prism two beams of parallel light emerge from the hypotenuse face, the angle between them being 6° . Calculate the value of angle B . You may assume that for small angles $\sin i/\sin r$ equals i/r .

How would you discover whether the angle is more or less than 90° ? (O. & C.)

13. Give a detailed description of the optical system of the compound microscope, explaining the problems which arise in the design of an object lens for a microscope.

A compound microscope has lenses of focal length 1 cm and 3 cm. An object is placed 1.2 cm from the object lens; if a virtual image is formed 25 cm from the eye, calculate the separation of the lenses and the magnification of the instrument. (O. & C.)

14. A projection lantern contains a condensing lens and a projection lens. Show clearly in a ray diagram the function of these lenses.

A lantern has a projection lens of focal length 25 cm and is required to be able to function when the distance from lantern to screen may vary from 6 m to 12 m. What range of movement for the lens must be provided in the focusing arrangement? What is the approximate value of the ratio of the magnifications at the two extreme distances? (W.)

15. A converging lens of focal length 20 cm and a diverging lens of focal length 10 cm are arranged for use as an opera glass. Draw a ray diagram to scale showing how the final image at infinity is produced, describing briefly how you do this, and derive the magnifying power.

When an object is placed 60 cm in front of the converging lens and the lenses are separated by a distance x , a real image is formed 30 cm beyond the diverging lens. Calculate x . (C.)

16. An astronomical telescope consisting of an objective focal length 60 cm and an eyepiece of focal length 3 cm is focused on the moon so that the final image is formed at the minimum distance of distinct vision (25 cm) from the eyepiece. Assuming that the diameter of the moon subtends an angle of $\frac{1}{2}^\circ$ at the objective, calculate (a) the angular magnification, (b) the actual size of the image seen.

How, with the same lenses, could an image of the moon, 10 cm in diameter, be formed on a photographic plate? (C.)

17. Explain, with the aid of a ray diagram, how a simple astronomical telescope employing two converging lenses may form an apparently enlarged image of a distant extended object. State with reasons where the eye should be placed to observe the image.

A telescope constructed from two converging lenses, one of focal length 250 cm, the other of focal length 2 cm, is used to observe a planet which subtends an angle of 5×10^{-5} radian. Explain how these lenses would be placed for normal adjustment and calculate the angle subtended at the eye of the observer by the final image.

How would you expect the performance of this telescope for observing a star to compare with one using a concave mirror as objective instead of a lens, assuming that the mirror had the same diameter and focal length as the lens. (O. & C.)

chapter twenty-four

Velocity of light Photometry

VELOCITY OF LIGHT

FOR many centuries the velocity of light was thought to be infinitely large; from about the end of the seventeenth century, however, evidence began to be obtained which showed that the speed of light, though enormous, was a finite quantity. Galileo, in 1600, attempted to measure the velocity of light by covering and uncovering a lantern at night, and timing how long the light took to reach an observer a few miles away. Owing to the enormous speed of light, however, the time was too small to measure, and the experiment was a failure. The first successful attempt to measure the velocity of light was made by RÖMER, a Danish astronomer, in 1676.

Römer's Astronomical Method

Römer was engaged in recording the eclipses of one of Jupiter's satellites or moons, which has a period of 1.77 days round Jupiter. The period of the satellite is thus very small compared with the period of the earth round the sun (one year), and the eclipses of the satellite occur very frequently while the earth moves only a very small distance in its orbit. Thus the eclipses may be regarded as *signals sent out from Jupiter* at comparatively short intervals, and observed on the earth; almost like a bright lamp covered at regular intervals at night and viewed by a distant observer.

The earth makes a complete revolution round the sun, S, in one year. Jupiter makes a complete revolution round the sun in about $11\frac{1}{2}$ years, and we shall assume for simplicity that the orbits of the earth and Jupiter are both circular, Fig. 24.1. At some time, Jupiter (J_1) and the earth (E_1) are on the same side of the sun, S, and in line with each other, and the earth and Jupiter are then said to be in *conjunction**. Suppose that an eclipse, or "signal", is now observed on the earth E_1 . If $E_1J_1 = x$, and c is the velocity of light, the time taken for the "signal" to reach E_1 is x/c ; and if the actual time when the eclipse occurred was a (which is not known), the time T_1 of the eclipse recorded on the earth is given by

$$T_1 = a + \frac{x}{c} \quad . \quad . \quad . \quad . \quad (i)$$

The earth and Jupiter now move round their respective orbits, and

* Astronomers use the terms 'conjunction' and 'opposition' in relation to the positions of the sun and Jupiter. With this usage, the latter are in 'opposition' here.

at some time, about $6\frac{1}{2}$ months later, the earth (E_2) and Jupiter (J_2) are on opposite sides of the sun S and in line with each other, Fig. 24.1. The earth and Jupiter are then said to be in *opposition*. During the $6\frac{1}{2}$ months

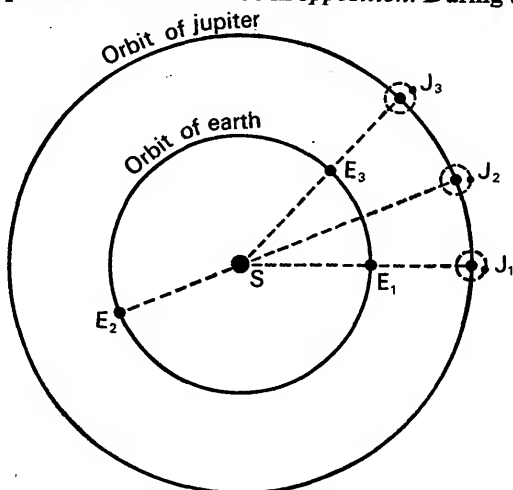


FIG. 24.1. Römer's method.

suppose that m eclipses have occurred at regular intervals T , i.e., T is the actual time between successive eclipses: the time for the interval between the 1st and m th eclipses is then $(m - 1) T$. In the position J_2, E_2 , however, the light travels a distance J_2E_2 , or $(x + d)$, from Jupiter to the earth, where d is the diameter of the earth's orbit round the sun. The time taken to travel this distance $= (x + d)/c$. Thus the time T' recorded on the earth at E_2 when the m th eclipse occurs is given by

$$T' = a + (m - 1) T + \frac{x + d}{c} \quad \text{. (ii)}$$

But, from (i), $T_1 = a + \frac{x}{c}$

Subtracting, $T' - T_1 = I = (m - 1) T + \frac{d}{c}$, (iii)

where I is the interval recorded on the earth for the time of m eclipses, from the position of conjunction of Jupiter and the earth to their position of opposition. By similar reasoning to the above, the interval I_1 recorded on the earth for m eclipses from the position of opposition (J_2, E_2) to the next position of conjunction of (J_3, E_3) is given by

$$I_1 = (m - 1) T - \frac{d}{c} \quad \text{. (iv)}$$

The reason why I_1 is less than I is that the earth is moving towards Jupiter from E_2 to E_3 , and away from Jupiter from E_1 to E_2 .

Römer observed that the m th eclipse between the position J_1, E_1 to the position J_2, E_2 occurred later than he expected by about $16\frac{1}{2}$ minutes; and he correctly deduced that the additional time was due to the time

taken by light to travel across the earth's orbit. In (iii), $(m - 1) T$ was the time expected for $(m - 1)$ eclipses, and d/c was the extra time ($16\frac{1}{2}$ minutes) recorded on the earth. Since $d = 300\,000\,000$ km approximately, the velocity of light, c , $= 300\,000\,000 / (16\frac{1}{2} \times 60) = 300\,000$ km per second approximately.

Römer also recorded that the time I for m eclipses from the position E_1, J_1 to the position E_2, J_2 was about 33 minutes more than the time I_1 , for m eclipses between position E_2, J_2 to the position E_3, J_3 . But, subtracting (iv) from (iii),

$$I - I_1 = \frac{2d}{c} \quad (1)$$

$$\therefore \frac{2d}{c} = 33 \text{ mins} = 33 \times 60 \text{ secs}$$

$$\therefore c = 2d/(33 \times 60) = 2 \times 300\,000\,000/(33 \times 60)$$

$$\therefore c = 300.000 \text{ km per second (approx.)}$$

Maximum and minimum observed periods. When the earth E_1 is moving directly away from Jupiter at J_a , the apparent period T' of the satellite is a maximum. Suppose the earth moves from E_1 to E_2 in the time T' , Fig. 24.2.

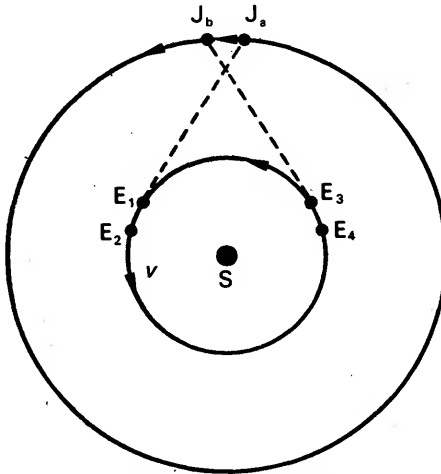


FIG. 24.2. Maximum and minimum periods.

Then if T is the actual period, v is the velocity of the earth, and c is the velocity of light, it follows that

$$T' = T + \frac{E_1 E_2}{c} = T + \frac{v T'}{c}$$

$$\therefore T' \left(1 - \frac{v}{c}\right) = T \quad \dots \quad (i)$$

When the earth is moving directly towards Jupiter at J_b , the apparent period T'' is a minimum. Suppose the earth moves from E_3 to E_4 in this time. Then

$$T'' = T - \frac{E_3 E_4}{c} = T - \frac{vT''}{c}$$

$$\therefore T'' \left(1 + \frac{v}{c}\right) = T \quad \quad \quad (ii)$$

From (i) and (ii), it follows that

$$T' \left(1 - \frac{v}{c}\right) = T'' \left(1 + \frac{v}{c}\right)$$

$$\therefore \frac{T'}{T''} = \frac{1 + \frac{v}{c}}{1 - \frac{v}{c}} = \text{ratio of maximum and minimum observed periods}$$

Bradley's Aberration Method

Römer's conclusions about the velocity of light were ignored by the scientists of his time. In 1729, however, the astronomer BRADLEY observed that the angular elevation of a "fixed" star varied slightly according to the position of the earth in its orbit round the sun. For some time he was puzzled by the observation. But while he was being rowed across a stream one day, he noticed that the boat drifted slightly downstream; and he saw immediately that the difference between the actual and observed angular elevation of the star was due to a combination of the velocity of the earth in its orbit (analogous to the velocity of the stream) with that of the velocity of light (analogous to the velocity of the boat). Thus if the earth were stationary, a telescope T would have to point in the true direction AS of a star S to observe it; but since the earth is moving in its orbit round the sun with a velocity v , T would have to be directed along MN to observe the star, where MN makes a small angle α with the direction SA , Fig. 24.3 (i).

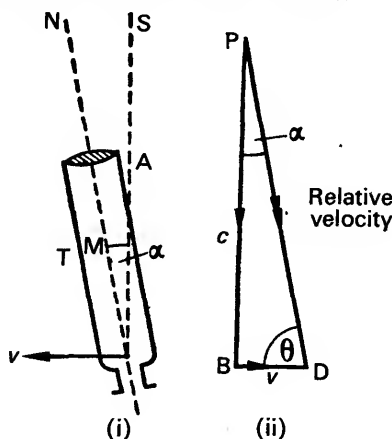


FIG. 24.3. Bradley's aberration method.

The direction of MN is that of the *relative velocity* between the earth and the light from S, which is found by subtracting the velocity v from the velocity c of the light. This is easily done by drawing the triangle of velocities PBD, in which PB, parallel to SA, represents c in magnitude and direction, while BD represents a velocity equal and opposite to v , Fig. 24.3 (ii). The resultant of PB and BD is then PD, which is parallel to NM, and PD represents the relative velocity.

The angle α between the true and apparent directions of the star is known as the *aberration*. From Fig. 24.3 (ii), it follows that

$$\frac{v}{c} = \frac{\sin \alpha}{\sin \theta},$$

where θ is the apparent altitude of the star.

$$\therefore c = \frac{v \sin \theta}{\sin \alpha} \quad (2)$$

Since α is very small, $\sin \alpha$ is equal to α in radians. By using known values of v , θ , and α , Bradley calculated c , the velocity of light, and obtained a value close to Römer's value. The aberration α is given by half the difference between the maximum and minimum values of the apparent altitude, θ , of the star.

Fizeau's Rotating Wheel Method. A Terrestrial Method

In 1849 FIZEAU succeeded in measuring the velocity of light with apparatus on the earth, for the first time. His method, unlike Römer's and Bradley's method, is thus known as a *terrestrial method*.

Fizeau's apparatus is illustrated in Fig. 24.4. A bright source at O emits light which is converged to a point H by means of the lens and the plane sheet of glass F, and is then incident on a lens B. H is at the focus of B, and the light thus emerges parallel after refraction through the latter and travels several miles to another lens C. This lens brings the light to a focus at M, where a silvered plane mirror is positioned, and the light is now returned back along its original path to the glass plate F. An image of O can thus be observed through F by a lens E.

The rim of a toothed wheel, W, which can rotate about a horizontal axis Q, is placed at H, and is the important feature of Fizeau's method.

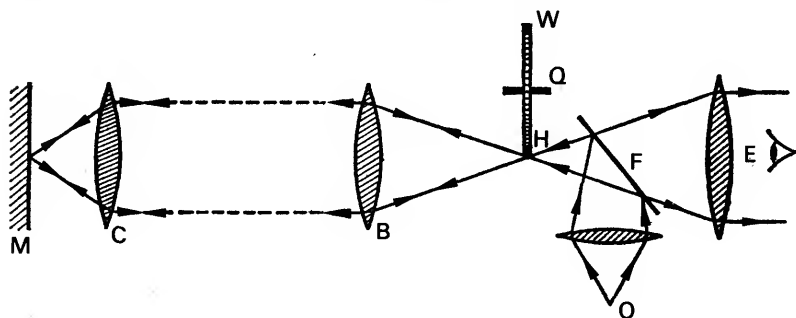


FIG. 24.4. Fizeau's rotating wheel method.

The teeth and the gaps of W have the same width, Fig. 24.5. As W is rotated, an image is observed through E as long as the light passes through the wheel towards E. When the speed of rotation exceeds about 10 cycles per second, the succession of images on the retina causes an image of O to be seen continuously. As the speed of W is further increased, however, a condition is reached when the returning light passing through a gap of W and reflected from M arrives back at the wheel to find that the tooth *next* to the gap has taken the place of the gap. Assuming the wheel is now driven at a constant speed, it can be seen that light continues to pass through a gap in the wheel towards M but always arrives back at W to find its path barred by the neighbouring tooth. The field of view through E is thus dark. If the speed of W is now doubled a bright field of view is again observed, as the light passing through a gap arrives back from M to find the next *gap* in its place, instead of a tooth as before.

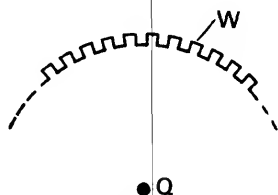


FIG. 24.5.
Rotating wheel.

Fizeau used a wheel with 720 teeth, and first obtained a dark field of view through E when the rate of revolution was 12.6 revs per second. The distance from H to M was 8633 metres. Thus the time taken for the light to travel from H to M and back = $2 \times 8633/c$ seconds, where c is the velocity of light in metres per second. But this is the time taken by a tooth to move to a position corresponding to a neighbouring gap. Since there are 2×720 teeth and gaps together, the time = $1/(2 \times 720)$ of the time taken to make one revolution, or $1/(2 \times 720 \times 12.6)$ seconds, as 12.6 revs are made in one second.

$$\therefore \frac{2 \times 8633}{c} = \frac{1}{2 \times 720 \times 12.6}$$

$$\therefore c = 2 \times 8633 \times 2 \times 720 \times 12.6$$

$$= 3.1 \times 10^8 \text{ metres per sec.}$$

The disadvantage of Fizeau's method is mainly that the field of view can never be made perfectly dark, owing to the light diffusely reflected at the teeth towards E. To overcome this disadvantage the teeth of the wheel were bevelled, but a new and more accurate method of determining the velocity of light was devised by FOUCAULT in 1862.

Foucault's Rotating Mirror Method

In Foucault's method a plane mirror M_1 is rotated at a high constant angular velocity about a vertical axis at A, Fig. 24.6. A lens L is placed so that light from a bright source at O_1 is reflected at M_1 and comes to a focus at a point P on a concave mirror C. The centre of curvature of C is at A, and consequently the light is reflected back from C along its original path, giving rise to an image coincident with O_1 . In order to observe the image, a plate of glass G is placed at 45° to the axis of the lens, from which the light is reflected to form an image at B_1 .

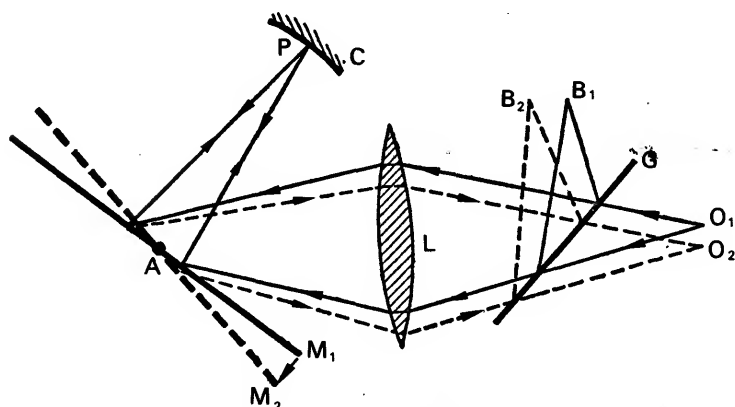


FIG. 24.6. Foucault's rotating mirror method.

Suppose the plane mirror M_1 begins to rotate. The light reflected by it is then incident on C for a fraction of a revolution, and if the speed of rotation is 2 revs per sec, an intermittent image is seen. As the speed of M_1 is increased to about 10 revs per sec the image is seen continuously as a result of the rapid impressions on the retina. As the speed is increased further, the light reflected from the mirror flashes across from M_1 to C , and returns to M_1 to find it displaced by a very small angle θ to a new position M_2 . An image is now observed at B_2 , and by measuring the displacement, B_1B_2 , of the image Foucault was able to calculate the velocity of light.

Theory of Foucault's Method

Consider the point P on the curved mirror from which the light is always reflected back to the plane mirror, Fig. 24.7. When the plane mirror is at M_1 , the image of P in it is at I_1 , where $AI_1 = AP = a$, the radius of curvature of C (see p. 396). The rays incident on the lens L from the plane mirror appear to come from I_1 . When the mirror is at M_2 the image of P in it is at I_2 , where $AI_2 = AP = a$, and the rays incident on L from the mirror now appear to come from I_2 . Now the

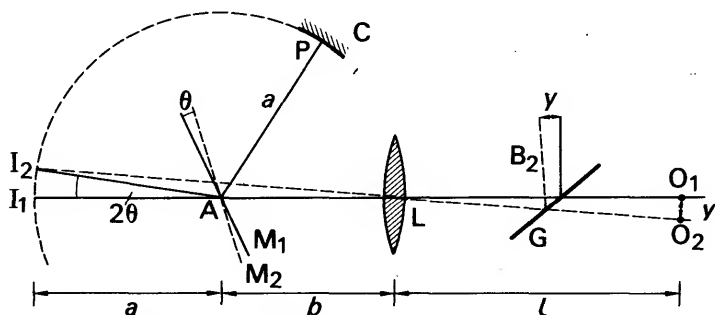


FIG. 24.7. Explanation of Foucault's method.

mirror has rotated through an angle θ from M_1 to M_2 , and the direction PA of the light incident on it is constant. The angle between the reflected rays is thus 2θ (see p. 393), and hence $I_2AI_1 = 2\theta$.

$$\therefore I_2I_1 = a \times 2\theta = 2a\theta \quad (i)$$

The images O_1, O_2 formed by the lens, L, are the images of I_1, I_2 in it, as the light incident on L from the mirror appear to come from I_1, I_2 . Hence $I_2I_1 : O_1O_2 = I_1L : LO_1$.

$$\therefore \frac{I_2I_1}{y} = \frac{(a+b)}{l}$$

where $y = O_1O_2$, $AL = b$, and $LO_1 = l$.

$$\therefore I_2I_1 = \frac{(a+b)y}{l} \quad (ii)$$

From (i) and (ii), it follows that

$$2a\theta = \frac{(a+b)y}{l}$$

$$\therefore \theta = \frac{(a+b)y}{2al} \quad (iii)$$

The angle θ can also be expressed in terms of the velocity of light, c , and the number of revolutions per second, m , of the plane mirror. The angular velocity of the mirror is $2\pi m$ radians per second, and hence the time taken to rotate through an angle θ radians is $\theta/2\pi m$ secs. But this is the time taken by the light to travel from the mirror to C and back, which is $2a/c$ secs.

$$\therefore \frac{\theta}{2\pi m} = \frac{2a}{c}$$

$$\therefore \theta = \frac{4\pi ma}{c} \quad (iv)$$

From (iii) and (iv), we have

$$\frac{(a+b)y}{2al} = \frac{4\pi ma}{c}$$

$$\therefore c = \frac{8\pi ma^2l}{(a+b)y} \quad (3)$$

As m, a, l, b are known, and the displacement $y = O_1O_2 = B_1B_2$ and can be measured, the velocity of light c can be measured.

The disadvantage of Foucault's method is mainly that the image obtained is not very bright, making observation difficult. Michelson (p. 559) increased the brightness of the image by placing a large lens between the plane mirror and C, so that light was incident on C for a greater fraction of the mirror's revolution. Since the distance a was increased at the same time, Fig. 24.7, the displacement of the image was also increased.

The velocity of light in water was observed by Foucault, who placed a pipe of water between the plane mirror and C. He found that, with

the number of revolutions per second of the mirror the same as when air was used, the displacement y of the image B was *greater*. Since the velocity of light $= 8\pi ma^2l/(a+b)y$, from (3), it follows that the velocity of light in water is *less* than in air. Newton's "corpuscular theory" of light predicted that light should travel faster in water than in air (p. 680), whereas the "wave theory" of light predicted that light should travel slower in water than in air. The direct observation of the velocity of light in water by Foucault's method showed that the corpuscular theory of Newton could not be true.

Michelson's Method for the Velocity of Light

The velocity of light, c , is a quantity which appears in many fundamental formulæ in advanced Physics, especially in connection with the theories concerning particles in atoms and calculations on atomic (nuclear) energy. EINSTEIN has shown, for example, that the energy W released from an atom is given by $W = mc^2$ joules, where m is the decrease in mass of the atom in kilogrammes and c the velocity of light in metres per second. A knowledge of the magnitude of c is thus important. A. A. MICHELSON, an American physicist, spent many years of his life in measuring the velocity of light, and the method he devised is regarded as one of the most accurate.

The essential features of Michelson's apparatus are shown in Fig. 24.8. X is an equiangular octagonal steel prism which can be rotated at constant speed about a vertical axis through its centre. The faces of the prism are highly polished, and the light passing through a slit from a very bright source O is reflected at the surface A towards a plane mirror B. From B the light is reflected to a plane mirror L, which is placed so that the image of O formed by this plane mirror is at the focus of a large concave mirror HD. The light then travels as a parallel

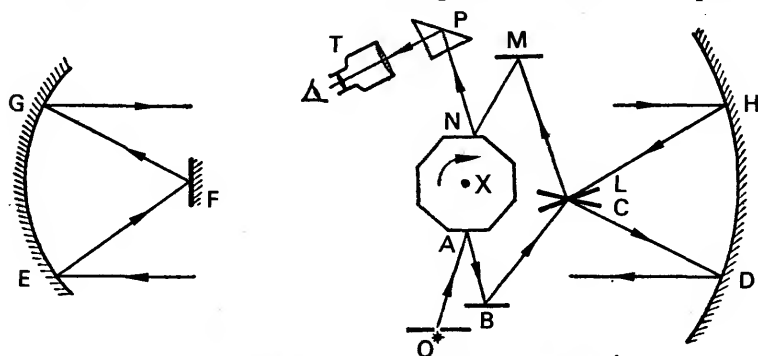


FIG. 24.8. Michelson's rotating prism method.

beam to another concave mirror GE a long distance away, and it is reflected to a plane mirror F at the focus of GE. The light is then reflected by the mirror, travels back to H, and is there reflected to a plane mirror C placed just below L and inclined to it as shown. From C the light is reflected to a plane mirror M, and is then incident on the face N of the octagonal prism opposite to A.

The final image thus obtained is viewed through T with the aid of a totally reflecting prism P.

The image is seen by light reflected from the top surface of the octagonal prism X. When the latter is rotated the image disappears at first, as the light reflected from A when the prism is just in the position shown in Fig. 24.8 arrives at the opposite face to find this surface in some position inclined to that shown. When the speed of rotation is increased and suitably adjusted, however, the image reappears and is seen in the same position as when the prism X is at rest. *The light reflected from A now arrives at the opposite surface in the time taken for the prism to rotate through 45° , or $\frac{1}{8}$ th of a revolution*, as in this case the surface on the left of N, for example, will occupy the latter's position when the light arrives at the upper surface of X.

Suppose d is the total distance in metres travelled by the light in its journey from A to the opposite face; the time taken is then d/c , where c is the velocity of light. But this is the time taken by X to make $\frac{1}{8}$ th of a revolution, which is $1/8m$ secs if the number of revolutions per second is m .

$$\therefore \frac{1}{8m} = \frac{d}{c}$$

$$\therefore c = 8md \text{ metres per sec.} \quad (4)$$

Thus c can be calculated from a knowledge of m and d .

Michelson performed the experiment in 1926, and again in 1931, when the light path was enclosed in an evacuated tube 1.6 km long. Multiple reflections were obtained to increase the effective path of the light. A prism with 32 faces was also used, and Michelson's result for the velocity of light *in vacuo* was 299 774 kilometres per second. Michelson died in 1931 while he was engaged in another measurement of the velocity of light.

EXAMPLES

1. Describe carefully Fizeau's method of determining the speed of propagation of light by means of a toothed wheel. Given that the distance of the mirror is 8000 m, that the revolving disc has 720 teeth, and that the first eclipse occurs when the angular velocity of the disc is $13\frac{1}{4}$ revolutions per second, calculate the speed of propagation of light. (*W*.)

First part. See text.

Second part. Suppose c is the speed of light in metres per second.

$$\therefore \text{time to travel to mirror and back} = \frac{2 \times 8000}{c} \text{ s.}$$

$$\text{But time for one tooth to occupy the next gap's position} = \frac{1}{13\frac{1}{4}} \times \frac{1}{2 \times 720} \text{ s.}$$

$$\therefore \frac{2 \times 8000}{c} = \frac{1}{13\frac{1}{4} \times 2 \times 720}$$

$$\therefore c = 2 \times 8000 \times 13\frac{1}{4} \times 2 \times 720 \\ = 3.2 \times 10^8 \text{ m s}^{-1}.$$

2. A beam of light is reflected by a rotating mirror on to a fixed mirror, which sends it back to the rotating mirror from which it is again reflected, and then makes an angle of 18° with its original direction. The distance between the two mirrors is 10^4 m, and the rotating mirror is making 375 revolutions per sec. Calculate the velocity of light. (*L.*)

Suppose OA is the original direction of the light, incident at A on the mirror in the position M_1 , B is the fixed mirror, and AC is the direction of the light reflected from the rotating mirror when it reaches the position M_2 , Fig. 24.9.

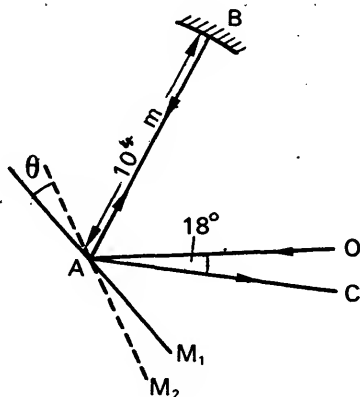


FIG. 24.9. Example.

The angle θ between M_1 , M_2 is $\frac{1}{2} \times 18^\circ$, since the angle of rotation of a mirror is half the angle of deviation of the reflected ray when the incident ray (BA in this case) is kept constant. Thus $\theta = 9^\circ$.

$$\text{Time taken by mirror to rotate } 360^\circ = \frac{1}{375} \text{ s.}$$

$$\therefore \text{ time taken to rotate } 9^\circ = \frac{9}{360} \times \frac{1}{375} \text{ s.}$$

But this is also the time taken by the light to travel from A to B and back, which is given by $2 \times 10^4/c$, where c is the velocity of light in m s^{-1} .

$$\therefore \frac{2 \times 10^4}{c} = \frac{9}{360} \times \frac{1}{375}$$

$$\therefore c = \frac{2 \times 10^4 \times 360 \times 375}{1} = 3 \times 10^8 \text{ m s}^{-1}.$$

Photometry

Standard Candle. The Candela

Light is a form of energy which stimulates the sensation of vision. The sun emits a continuous stream of energy, consisting of ultra-violet, visible, and infra-red radiations (p. 344), all of which enter the eye; but only the energy in the visible radiations, which is called *luminous energy*, stimulates the sensation of vision. In Photometry we are concerned only with the luminous energy emitted by a source of light.

Years ago the luminous energy per second from a candle of specified

wax material and wick was used as a standard of luminous energy. This was called the *British Standard Candle*. The luminous energy per second from any other source of light was reckoned in terms of the standard candle, and its value was given at 10 candle-power (10 c.p.) for example. As the standard candle was difficult to reproduce exactly, the standard was altered. It was defined as one-tenth of the intensity of the flame of the *Vernon Harcourt pentane lamp*, which burns a mixture of air and pentane vapour under specified conditions. Later it was agreed to use as a standard the *international standard candle*, which is defined in terms of the luminous energy per second from a particular electric lamp filament maintained under specified conditions, but the precision of this standard was found to be unsatisfactory. In 1948, a unit known as the *candela*, symbol "cd", was adopted. This is defined as the luminous intensity of $1/600\,000$ metre² ($1/60$ cm²) of the surface of a black body at the temperature of freezing platinum under 101 325 newtons per metre² pressure. A standard is maintained at the National Physical Laboratory.

Illumination and its Units

If a lamp S of 1 candela is placed 1 metre away from a small area A and directly in front of it, the *illumination* of the surface of A is said to be 1 *metre-candle* or *lux*, Fig. 24.10. If the same lamp is placed 1 centimetre away from A, instead of 1 m, the illumination of the surface is said to be 1 *cm-candle* (or 1 *phot*). The SI unit of illumination is the lux (see also p. 564). The "foot-candle" has been used as a unit; the distance of 1 metre in Fig. 24.10 is replaced by 1 foot. $1 \text{ lux} = 10^{-4} \text{ phot} = 9.3 \times 10^{-2} \text{ foot-candle}$. It is recommended that offices should have an intensity of illumination of about 90 lux, and that the intensity of illumination for sewing dark materials in workrooms should be about 200 lux.

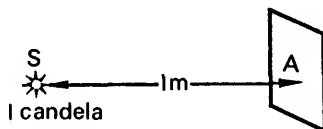


FIG. 24.10. Metre-candle (lux).

Luminous Flux, F

In practice a source of light emits a continuous stream of energy, and the name *luminous flux* has been given to the *luminous energy emitted per second*. The unit of luminous flux is the **lumen**, lm. Since a lumen is a certain amount of "energy per second", or "power", there must be a relation between the lumen and the watt, the mechanical unit of power; and experiment shows that 621 lumens of a green light of wavelength 5.540×10^{-10} m is equivalent to 1 watt.

Solid Angle

A lamp radiates luminous flux in all directions round it. If we think of a particular small lamp and a certain direction from it, for example that of the corner of a table, we can see that the flux is radiated towards the corner in a cone whose apex is the lamp. A thorough study of

where F is the luminous flux contained in the small cone. Thus the luminous intensity of the source is the *luminous flux per unit solid angle* in the particular direction. It can now be seen that "luminous intensity" is a measure of the "luminous flux density" in the direction concerned.

The unit of luminous intensity of a source is the *candela*, defined on p. 562, and the luminous intensity was formerly known as the *candle-power* of the source. When the luminous flux, F , in the cone in Fig. 24.12 is 1 lumen (the unit of luminous flux), and the solid angle, ω , of the cone is 1 unit, it follows from equation (6) that $I = 1$ candela. Thus the *lumen can be defined as the luminous flux radiated within unit solid angle by a uniform source of one candela*. A small source of I candela radiates $4\pi I$ lumens all round it.

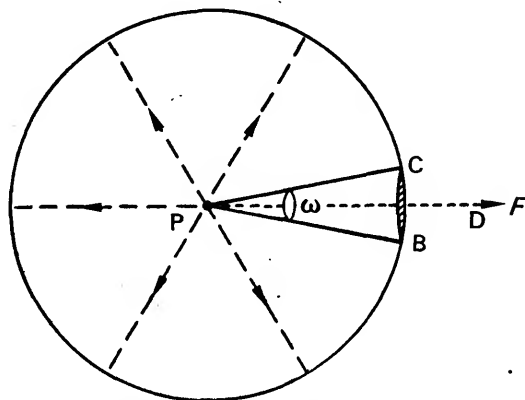


FIG. 24.12. Luminous intensity of source.

Illumination of Surface

On p. 562 we encountered various units of *illumination of a surface*; these were the metre-candle or the lux (lx), which is the SI unit, the cm-candle and the foot-candle. The illumination is defined generally as the *luminous flux per unit area* falling on the part of the surface under consideration. Thus if F is the luminous flux incident on a small area A , the intensity of illumination, E , is given by

$$E = \frac{F}{A} \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

When $F = 1$ lumen and $A = 1 \text{ m}^2$, then $E = 1 \text{ lm m}^{-2}$. Thus 1 m-candle of illumination is equivalent to an illumination of 1 lumen per m^2 of the surface. Similarly, 1 cm-candle (p. 562) is equivalent to an illumination of 1 lumen per cm^2 of the surface. $1 \text{ lux} = 1 \text{ lm m}^{-2}$.

Suppose a lamp of 50 candela illuminates an area of 2 m^2 at a distance 20 m away. The flux F emitted all round the lamp $= I\omega = I \times 4\pi = 50 \times 4\pi = 200\pi$ lumens. The flux per unit area at a distance of 20 m away is given by

$$\frac{200\pi}{4\pi r^2} = \frac{200\pi}{4\pi \times 20^2} \text{ lumens per m}^2$$

\therefore flux falling on an area of 2 m^2

$$= 2 \times \frac{200\pi}{4\pi \times 20^2} = 0.25 \text{ lumens.}$$

The reader should take pains to distinguish carefully between the meaning of "luminous intensity" and "illumination" and their units. The former refers to the *source* of light and is measured in candelas (formerly candle-power); the latter refers to the *surface* illuminated and is measured in lux or metre-candles. Further, "luminous intensity" is defined in terms of unit solid angle, which concerns three dimensions whereas "illumination" is defined in terms of unit area, which concerns two dimensions.

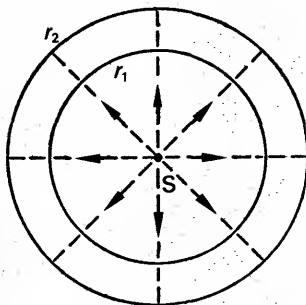
Relation between Luminous Intensity (I) and Illumination (E)

Consider a *point* source of light of uniform intensity and a small part X of a surface which it illuminates. If the source of light is doubled, the illumination of X is doubled because the luminous flux incident on it is twice as much. Thus

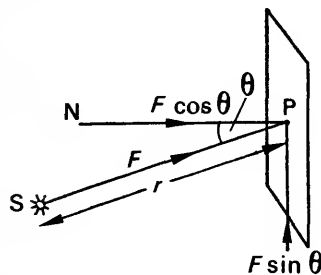
$$E \propto I \quad \dots \dots \dots (i)$$

where E is the illumination due to a point source of luminous intensity I at a given place.

The illumination of the surface also depends on the distance of X from the source. Suppose two spheres of radii r_1, r_2 are drawn round a point source of intensity I , such as S in Fig. 24.13 (i). The same amount



(i)



(ii)

FIG. 24.13. (i). Inverse square law. (ii). Lambert's cosine rule.

of luminous flux F spreads over the surface area ($4\pi r^2$) of both spheres, and hence $E_1 : E_2 = F/4\pi r_1^2$, where E_1, E_2 are the values of the illumination at the surface of the smaller and larger spheres respectively. Thus $E_1 : E_2 = r_2^2 : r_1^2$. It can hence be seen that the illumination due to a given point source varies inversely as the square of the distance from the source, i.e.,

$$E \propto \frac{1}{r^2} \quad \dots \dots \dots (ii)$$

In the eighteenth century LAMBERT showed that the illumination round a particular point P on a surface is proportional to $\cos \theta$, where θ is the angle between the normal PN at P to the surface and the line SP joining the source S to P, Fig. 24.13 (ii). A rigid proof of Lambert's law is given shortly, but we can easily see qualitatively why the cosine rule is true. The luminous flux F illuminating P is in the direction SP, and thus has a component $F \cos \theta$ along NP and a component $F \sin \theta$ parallel to the surface, Fig. 24.13. The latter does not illuminate the surface. Hence the effective part of the flux F is $F \cos \theta$.

From equations (i) and (ii) and the cosine law, it follows that the illumination E round P due to the source S of luminous intensity I is given by

$$E = \frac{I \cos \theta}{r^2} \quad \dots \dots \dots (8)$$

where $SP = r$. This is a fundamental equation in Photometry, and it is proved rigidly on p. 567. In applying it in practice one has to take into account that (i) a "point source" is difficult to realise, (ii) the area round the point considered on the surface should be very small so that the flux incident all over it can be considered the same, (iii) the intensity I of a source varies in different directions, (iv) the actual value of illumination round a point on a table, for example, is not only due to the electric lamp above it but also to the luminous flux diffusely reflected towards the point from neighbouring objects such as walls.

Example: Suppose that we are required to calculate the intensity I of a lamp S fixed 4 m above a horizontal table, if the value of the illumination at a point P on the table 3 m to one side of the vertical through the lamp is 6 m-candles, Fig. 24.14.

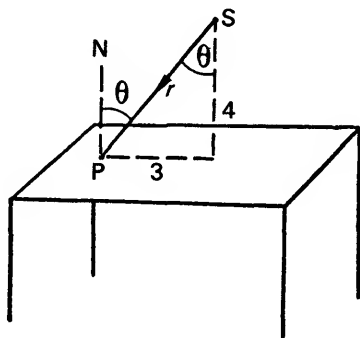


FIG. 24.14. Example.

The illumination, E , at P is given by

$$E = \frac{I \cos \theta}{r^2} \quad \dots \dots \dots (i)$$

where I is the luminous intensity of S, θ is the angle between SP and the normal PN to the table at P, and $r = SP$.

But

$$r^2 = 4^2 + 3^2 = 25, \text{ i.e., } r = 5,$$

$$\cos \theta = \frac{4}{r} = \frac{4}{5},$$

and

$$E = 6 \text{ m-candles.}$$

Substituting in (i),

$$\therefore 6 = \frac{I \times \frac{4}{5}}{25}$$

$$\therefore I = 187.5 \text{ candelas}$$

Proof of $E = I \cos \theta / r^2$. Consider a source S of intensity I illuminating a very small area A round a point P on a surface, Fig. 24.15. If ω is the solid angle at S in the cone obtained by joining S to the boundary of the area, the illumination at P is given by

$$E = \frac{F}{A} = \frac{I\omega}{A}$$

as $I = \frac{F}{\omega}$ (p. 563). Now, by definition,

$$\omega = \frac{A_1}{r^2},$$

where A_1 is the area cut off on a sphere of centre S and radius r ($= SP$) by the generating lines of the cone.

$$\therefore E = \frac{I\omega}{A} = \frac{I A_1}{r^2 A},$$

But

$$A_1 = A \cos \theta,$$

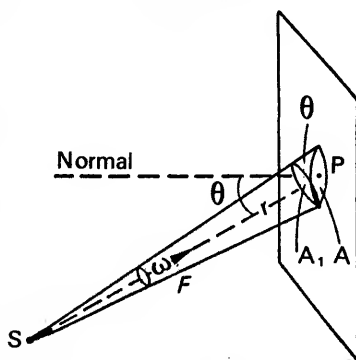


FIG. 24.15. Proof of $E = I \cos \theta / r^2$.

since A_1 is the projection of A on the sphere of centre S , and θ is the angle between the two areas as well as the angle made by SP with the normal to the surface, Fig. 24.15.

$$\therefore E = \frac{I A \cos \theta}{r^2 A} = \frac{I \cos \theta}{r^2}$$

Luminance of a Surface: Reflection and Transmission Factors

The *luminance* of a surface in a given direction is defined as the luminous flux per unit area *coming from* the surface in the particular direction. The luminance of white paint on the wall of a room is considerably higher than the luminance of a brown-painted panel in the middle of the wall; the luminance of a steel nib is much greater than that of a dark ebonite penholder.

The "luminance" of a particular surface should be carefully distinguished from the "illumination" of the surface, which is the luminous flux per sq m *incident on* the surface. Thus the illumination of white chalk on a particular blackboard is practically the same as that of the neighbouring points on the board itself, whereas the luminance of the chalk is much greater than that of the board. The difference in luminance

Lamps of luminous intensities I_1 and I_2 respectively are placed on opposite sides of a white opaque screen, and some of the diffusely-reflected light from the opposite surfaces A, B is incident on two identical totally reflecting prisms P, Q, Fig. 24.16 (i). The light reflected from the prisms then passes towards the "Lummer-Brodhun cube", which is the main feature of the photometer. This consists of two right-angled isosceles prisms in optical contact at their central portion C, but with the edges of one cut away so that an air-film exists at M, N all round C between the prisms. The rays leaving the prism P are thus transmitted through the central portion C of the "cube", but totally reflected at the edges. Similarly, the rays reflected from the prism Q towards the "cube" are transmitted through the central portion but totally reflected at the edges. An observer of the "cube" thus sees a central circular patch b of light due initially to the light from the source of intensity I_1 , and an outer portion a due initially to the light from the source of intensity I_2 Fig. 24.16 (ii).

Comparison of Luminous Intensities

In general, the brightness of the central and outer portions of the field of view in a Lummer-Brodhun photometer is different, so that one appears darker than the other. By moving one of the sources, however, a position is obtained when both portions appear equally bright, in which case they cannot be distinguished from each other and the field of view is uniformly bright. A "photometric balance" is then said to exist.

Suppose that the distances of the sources I_1 and I from the screen are d_1 , d respectively, Fig. 24.16. The intensity of illumination, E , due to the source I is generally given by $E = I \cos \theta / d^2$ (p. 566). But $\theta = 0$ in this case, as the line joining the source to the screen is normal to the screen. Hence, since $\cos 0^\circ = 1$, $E = I/d^2$. Similarly, the intensity of illumination, E_1 , at the screen due to the source I_1 is given by $E_1 = I_1/d_1^2$. Now the luminance of the surface B $= r_1 E_1$, where r_1 is the reflection factor of the surface (p. 568); and the luminance of the surface A $= rE$, where r is the reflection factor of this surface. Hence, for a photometric balance,

$$\begin{aligned} r_1 E_1 &= rE \\ \therefore \frac{r_1 I_1}{d_1^2} &= \frac{rI}{d^2} \end{aligned} \quad (i)$$

If the reflection factors r_1 , r of A, B are equal, equation (i) becomes

$$\begin{aligned} \frac{I_1}{d_1^2} &= \frac{I}{d^2} \\ \therefore \frac{I_1}{I} &= \frac{d_1^2}{d^2} \end{aligned}$$

The ratio of the intensities are hence proportional to the squares of the corresponding distances of the sources from the screen.

The reflection factors r_1 , r are not likely to be exactly equal, however, in which case another or *auxiliary lamp* is required to compare the

candle-powers I_1 , I . The auxiliary lamp, I_2 , is placed on the right side, say, of the screen at a distance d_2 , and *one* of the other lamps is placed on the other side. A photometric balance is then obtained,

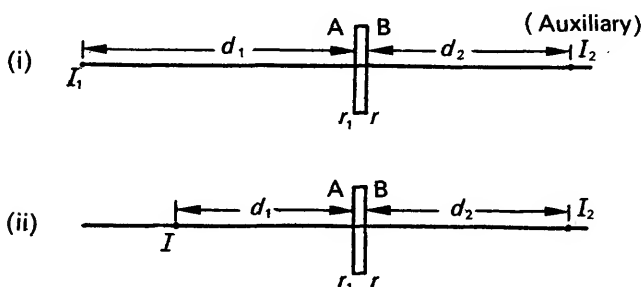


FIG. 24.17. Comparison of luminous intensities.

Fig. 24.17 (i). In this case, if I_1 is the intensity of the lamp and d_1 is its distance from the screen,

$$r_1 \frac{I_1}{d_1^2} = r \frac{I_2}{d_2^2} \quad \text{(ii)}$$

The remaining lamp I is then used instead of the lamp I_1 , and a photometric balance is again obtained by moving this lamp, *keeping the position of the lamp I_2 unaltered*, Fig. 24.17 (ii). Suppose the distance of the lamp I from the screen is d . Then

$$r_1 \frac{I}{d^2} = r \frac{I_2}{d_2^2} \quad \text{(iii)}$$

From (ii) and (iii), it follows that

$$\begin{aligned} r_1 \frac{I_1}{d_1^2} &= r_1 \frac{I}{d^2} \\ \therefore \frac{I_1}{d_1^2} &= \frac{I}{d^2} \\ \therefore \frac{I_1}{I} &= \frac{d_1^2}{d^2} \end{aligned} \quad \text{(13)}$$

The intensities I_1 , I are hence proportional to the squares of the corresponding lamp distances from the screen. It should be noted from (13) that the auxiliary lamp's intensity is not required in the comparison of I_1 and I , nor is its constant distance d_2 from the screen required.

Measurement of Illumination

It was pointed out at the beginning of the chapter that the maintenance of standards of illumination plays an important part in safeguarding our health. It is recommended that desks in class-rooms and

offices should have an illumination of 60–110 lux, and workshops an illumination of 120–180 lux; for sewing dark materials an intensity of 180–300 lux is recommended, while 1200 lux is suggested for the operating table in a hospital.

Photovoltaic cell. There are two types of meters for measuring the illumination of a particular surface. A modern type is the photovoltaic cell, which may consist of a cuprous oxide and copper plate, made by

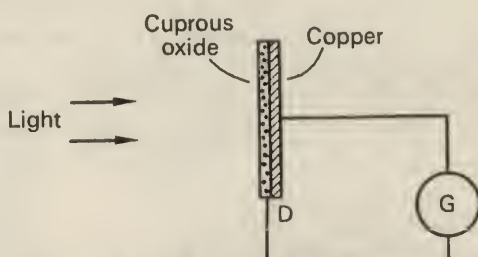


FIG. 24.18. Photovoltaic cell.

oxidising one side of a copper disc D, Fig. 24.18. When the oxide surface is illuminated, electrons are emitted from the surface whose number is proportional to the incident luminous energy, and a current flows in the microammeter or sensitive moving-coil galvanometer G which is proportional to the illumination. The galvanometer is previously calibrated by placing a standard lamp at known distances from the disc D, and its scale reads lux directly. Fig. 24.19 illustrates an "AVO Light-meter", which operates on this principle; it is simply laid on the surface whose illumination is required and the reading is then taken.



FIG. 24.19. AVO light-meter.

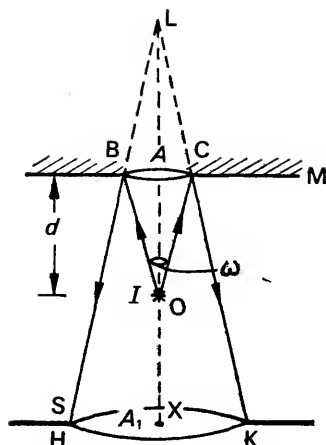


FIG. 24.20. Plane mirror illumination

Illumination due to plane mirror. Consider a small lamp of I candela at O in front of a plane mirror M , Fig. 24.20. The flux F from O in a small cone OBC of solid angle ω is $I\omega$, and falls on an area A of the mirror. This flux is reflected to illuminate an area A_1 , or HK , on a screen S in front of the mirror. Assuming the reflection factor of the mirror is unity,

$$\text{illumination of } S, E = \frac{F}{A_1} = \frac{I\omega}{A_1}.$$

But $\omega = \text{area } A/d^2$, where d is distance of O from the mirror.

$$\therefore E = \frac{IA}{d^2 A_1}.$$

But $A/A_1 = d^2/LX^2$, from similar triangles LBC , LHK .

$$\therefore E = \frac{I}{LX^2} \quad \dots \quad (i)$$

The image of O in the plane mirror is at L , which is at a distance LX from the screen S . It follows from (i) that, owing to light reflected from the mirror, the illumination of S is the same as that obtained by a lamp of I candela at the position of the image of O .

EXAMPLE

Define *lumen* and *lux*, and show how they are related. Describe and explain how you would make an accurate comparison of the illuminating powers of two lamps of the same type. A photometric balance is obtained between two lamps A and B when B is 100 cm from the photometer. When a block of glass G is placed between A and the photometer, balance is restored by moving B through 5 cm. Where must B be placed in order to maintain the balance when *two more* blocks, identical with G , are similarly placed between A and the photometer. (*L.*)

First part. See text.

Second part. Suppose I_1 , I_2 are the intensities of A , B respectively, and d is the distance of A from the photometer P . Then, originally,

$$\frac{I_1}{d^2} = \frac{I_2}{100^2} \quad \dots \quad (i)$$

When G is placed in position, the balance is restored by moving B 105 cm. from P , Fig. 24.21 (i). If t is the transmission factor of G , the effective intensity of A is tI_1 , and hence

$$\frac{tI_1}{d^2} = \frac{I_2}{105^2} \quad \dots \quad (ii)$$

Dividing (ii) by (i),

$$\therefore t = \frac{100^2}{105^2} \quad \dots \quad (iii)$$

When two more blocks are placed beside G, the effective intensity of $A = t \times t \times t I_1 = t^3 I_1$, Fig. 24.21 (ii). Thus if the distance of B from P is now x .

$$\frac{t^3 I_1}{d^2} = \frac{I_2}{x^2} \quad \dots \dots \dots (iv)$$

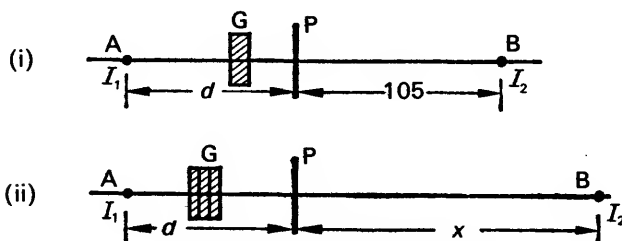


FIG. 24.21. Plane mirror illumination

Dividing (iv) by (i)

$$\therefore t^3 = \frac{100^2}{x^2}$$

From (iii)

$$\therefore \left(\frac{100^2}{105^2} \right)^3 = \frac{100^2}{x^2}$$

$$\therefore x^2 = \frac{100^2 \times 105^6}{100^6} = \frac{105^6}{100^4}$$

$$\therefore x = \frac{105^3}{100^2} = 116 \text{ cm.}$$

EXERCISES 24

Velocity of Light

1. In Fizeau's rotating wheel experiment the number of teeth was 720 and the distance between the wheel and reflector was 8633 metres. Calculate the number of revolutions per second of the wheel when extinction first occurs, assuming the velocity of light is $3 \cdot 13 \times 10^8$ metres per second. What are the disadvantages of Fizeau's method?

2. Draw a diagram of Foucault's method of measuring the velocity of light. How has the velocity of light in water been shown to be less than in air? The radius of curvature of the curved mirror is 20 metres and the plane mirror is rotated at 20 revs per second. Calculate the angle in degrees between a ray incident on the plane mirror and then reflected from it after the light has travelled to the curved mirror and back to the plane mirror (velocity of light = 3×10^8 m s⁻¹).

3. Draw a diagram showing the arrangement of the apparatus and the path of the rays of light in Fizeau's toothed wheel method for measuring the velocity of light. What are the chief difficulties met with in carrying out the experiment?

If the wheel has 150 teeth and 150 spaces of equal width and its distance from the mirror be 12 kilometres, at what speed, in revolutions per minute, will the first eclipse occur? (N .)

4. Explain how the velocity of light was first determined, and describe one more recent method of measuring it.

A beam of light after reflection at a plane mirror, rotating 2000 times per minute, passes to a distant reflector. It returns to the rotating mirror from which it is reflected to make an angle of 1° with its original direction. Assuming that the velocity of light is $300\,000\text{ km s}^{-1}$, calculate the distance between the mirrors. (*L.*)

5. Describe and explain in detail one accurate terrestrial method for measuring the speed of light in air.

How does the speed of light depend on (a) the wavelength of the light, (b) the medium through which it travels? (*O. & C.*)

6. Explain why the velocity of light is difficult to measure by a direct terrestrial method; illustrate your answer with an estimate of the orders of magnitude of the quantities involved in the assessment.

Describe, with the aid of a labelled diagram, Michelson's method for the determination of the velocity of light.

What are the advantages of this method over the earlier one developed by Foucault? (*N.*)

7. Describe any one terrestrial method by which the velocity of light has been determined. Why is a knowledge of the value of this velocity so important?

A beam of light can be interrupted $(1.0 \pm 0.0002) \times 10^7$ times each second on passing through a certain crystal device. Such a beam is reflected from a distant mirror and returned through the same crystal. It is found that for certain positions of the mirror very little light emerges from the crystal a second time. Explain this. One such position of the mirror is known to be between 18 and 26 metres from the crystal. Calculate a more accurate value for this distance, assuming the velocity of light to be $(3.0 \pm 0.002) \times 10^{10}\text{ cm s}^{-1}$. Estimate the error in this calculated value. (*C.*)

8. Describe a method of measuring the speed of light. Explain precisely what observations are made and how the speed is calculated from the experimental data.

A horizontal beam of light is reflected by a vertical plane mirror *A*, travels a distance of 250 metres, is then reflected back along the same path and is finally reflected again by the mirror *A*. When *A* is rotated with constant angular velocity about a vertical axis in its plane, the emergent beam is deviated through an angle of 18 minutes. Calculate the number of revolutions per second made by the mirror.

If an atom may be considered to radiate light of wavelength 5000 \AA for a time of 10^{-10} second, how many cycles does the emitted wave train contain? (*O. & C.*)

9. Give one reason why it is important to know accurately the velocity of light.

A glass prism whose cross-section is a regular polygon of n sides has its faces silvered and is mounted so that it can rotate about a vertical axis through the centre of the cross-section, with the silvered faces vertical. An intense narrow horizontal beam of light from a small source is reflected from one face of the prism to a small distant mirror and back to the same face, where it is reflected and used to form an image of the source which can be observed. Describe the behaviour of this image if the prism is set into rotation and slowly accelerated to a high speed. If the distance between the mirrors is D and the velocity of light is c , find an expression for the values of the angular velocity of the prism at which the image will be formed in its original position.

If $c = 3 \times 10^8\text{ m s}^{-1}$, $D = 30\text{ km}$ and the maximum safe speed for the mirror is 700 rev per second, what would be a suitable value for n ? (*O. & C.*)

10. Describe a terrestrial method by which the velocity of light has been measured. How could the method be modified to show that the velocity of light in water is less than that in air? Briefly discuss the theoretical importance of this fact. (C.)

11. Describe a terrestrial method of determining the velocity of light in air, explaining (without detailed calculation) how the result is obtained.

Why do we conclude that in free space red and blue light travel with the same speed, but that in glass red light travels faster than blue? (L.)

Photometry

12. A 30 candelas (cd) lamp X is 40 cm in front of a photometer screen. What is the illumination directly in front of X on the screen? Calculate the cd of a lamp Y which provides the same illumination when placed 60 cm from the screen.

13. A lamp of 800 cd is suspended 16 m above a road. Find the illumination on the road (i) at a point A directly below the lamp, (ii) at a point B 12 m from A.

14. Define the terms *luminous intensity* and *illumination*. Describe an accurate method of comparing the luminous intensities of two sources of light.

A lamp is fixed 4 m above a horizontal table. At a point on the table 3 m to one side of the vertical through the lamp, a light-meter is placed flat on the table. It registers 4 m candles. Calculate the intensity of the lamp. (C.)

15. What is meant by *luminous intensity* and *illumination*? How are they related to each other?

A small source of 32 cd giving out light equally in all directions is situated at the centre of a sphere of 8 m diameter, the inner surface of which is painted black. What is the illumination of the surface?

If the inner surface is repainted with a matt white paint which causes it to reflect diffusely 80 per cent of all lighting falling on it, what will the illumination be? (L.)

16. Describe an accurate form of photometer for comparing the luminous intensities of lamps.

A lamp is 100 cm from one side of a photometer and produces the same illumination as a second lamp placed at 120 cm on the opposite side. When a lightly smoked glass plate is placed before the weaker lamp, the brighter one has to be moved 50 cm to restore the equality of illumination. Find what fraction of the incident light is transmitted by the plate. (L.)

17. How would you compare the luminous intensities of two small lamps?

A small 100 cd lamp is placed 10 m above the centre of a horizontal rectangular table measuring 6 m by 4 m. What are the maximum and minimum values of the illumination on the table due to direct light?

How would your results be changed by the presence of a large horizontal mirror, placed 2 m above the lamp, so as to reflect light down on to the table, assuming that only 80 per cent of the light incident on the mirror is reflected? (W.)

18. Describe one form of a photometer, and explain how you would measure the light loss which results from enclosing a light source by a glass globe. Two small 16 cd lamps are placed on the same side of a screen at

distances of 2 and 5 m from it. Calculate the distance at which a single 32 cd lamp must be placed in order to give the same intensity of illumination on the screen. (N.)

19. Define *lumen*, *metre-candle*. Describe the construction and use of a Lummer-Brodhun photometer.

Twenty per cent of the light emitted by a source of 500 cd is evenly distributed over a circular area 5 m in diameter. What is the illumination at points within this area? (L.)

20. Distinguish between *luminous intensity* and *illumination*. How may the two sources be accurately compared?

A surface receives light normally from a source at a distance of 3 m. If the source is moved closer until the distance is only 2 m, through what angle must the surface be turned to reduce the illumination to its original value? (O. & C.)

21. Define *illumination of a screen*, *luminous intensity of a source of light*. Indicate units in which each of these quantities may be measured.

Describe a reliable photometer, and explain how you would use it to compare the reflecting powers of plaster of Paris and ground glass. (C.)

22. Describe an accurate form of photometer for comparing the luminous intensities of two sources of light.

Two electric lamps, A and B, are found to give equal illuminations on the two sides of a photometer when their distances from the photometer are in the ratio 4 : 5. A sheet of glass is then placed in front of B, and it is found that equality of illumination is obtained when the distances of A and B are in the ratio 16 : 19. Find the percentage of light transmitted by the glass. (C.)

WAVES AND SOUND

chapter twenty-five

Oscillations and Waves. Sound Waves

In this chapter we shall study the properties of oscillations and waves in general. Topics in waves which concern particular branches of the subject are discussed elsewhere in this book. We begin with a summary of the results relating to simple harmonic motion already derived (p. 45).

S.H.M.

Simple harmonic motion (S.H.M.) occurs when the force acting on an object or system is directly proportional to its displacement x from a fixed point and is always directed towards this point. If the object executes S.H.M., then the variation of the displacement x with time t can be written as

$$x = a \sin \omega t. \quad (1)$$

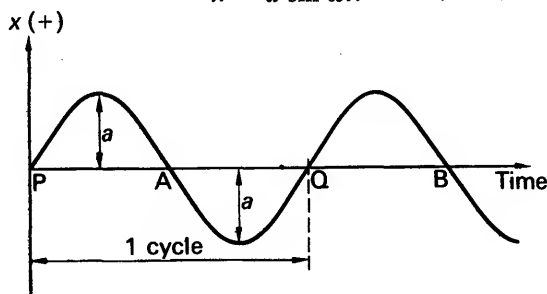


FIG. 25.1 Sine curve.

Here a is the greatest displacement from the mean or equilibrium position; a is the *amplitude*, Fig. 25.1. The constant ω is the 'angular frequency', and $\omega = 2\pi f$ where f is the *frequency* of vibration or number of cycles per second. The period T of the motion, or time to undergo one complete cycle is equal to $1/f$, so that $\omega = 2\pi/T$.

The small oscillation of a pendulum bob or vibrating layer of air is a *mechanical oscillation*, so that x is a displacement from a mean fixed position. Later, *electrical oscillations* are considered. x may then represent the instantaneous charge on the plates of a capacitor when the charge alternates about a mean value of zero. In an *electromagnetic wave*, x may represent the component of the electric or magnetic field vectors at a particular place.

Energy in S.H.M.

On p. 53, it was shown that the sum of the potential and kinetic energies of a body moving with S.H.M. is *constant* and equal to the total energy in the vibration. Further, it was shown that the time averages of the potential energy (P.E.) and kinetic energy (K.E.) are equal; each is half the total energy. In any mechanical oscillation, there is a continuous interchange or exchange of energy from P.E. to K.E. and back again.

For vibrations to occur, therefore, an agency is required which can possess and store P.E. and another which can possess and store K.E. This was the case for a mass oscillating on the end of a spring, as we saw on p. 50. The mass stores K.E. and the spring stores P.E.; and interchange occurs continuously from one to the other as the spring is compressed and released alternately. In the oscillations of a simple pendulum, the mass stores K.E. as it swings downwards from the end of an oscillation, and this is changed to P.E. as the height of the bob increases above its mean position.

Note that some agency is needed to accomplish the transfer of energy. In the case of the mass and spring, the force in the spring causes the transfer. In the case of the pendulum, the component of the weight along the arc of the circle causes the change from P.E. to K.E.

Electrical Oscillations

So far we have dealt with mechanical oscillations and energy. The energy in electrical oscillations takes a different form. There are still two types of energy. One is the energy stored in the electric field, and the other that stored in the magnetic field. To obtain electrical oscillations,

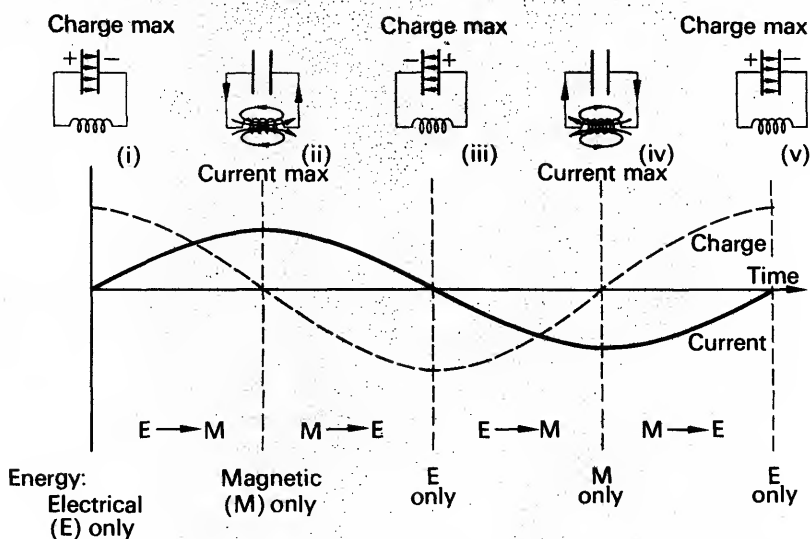


FIG. 25.2 Electrical oscillations – energy exchanges.

an inductor (coil) is used to produce the magnetic field and a capacitor to produce the electric field (see also p. 925).

Suppose the capacitor is charged and there is no current at this moment, Fig. 25.2 (i). A p.d. then exists across the capacitor and an electric field is present between the plates. At this instant all the energy is stored in the electric field, and since the current is zero there is no magnetic energy. Because of the p.d. a current will begin to flow and magnetic energy will begin to be stored in the inductor. Thus there will be a change from electric to magnetic energy. The p.d. is the agency which causes the transfer of energy.

One quarter of a cycle later the capacitor will be fully discharged and the current will be at its greatest, so that the energy is now entirely stored in the magnetic field, Fig. 25.2 (ii). The current continues to flow for a further quarter-cycle until the capacitor is fully charged in the opposite direction, when the energy is again completely stored in the electric field, Fig. 25.2 (iii). The current then reverses and the processes occur in reverse order, Fig. 25.2 (iv), after which the original state is restored and a complete oscillation has taken place, Fig. 25.2 (v). The whole process then repeats, giving continuous oscillations.

Phase of vibrations

Consider an oscillation given by $x_1 = a \sin \omega t$. Suppose a second oscillation has the same amplitude, a , and angular frequency, ω , but reaches the end of its oscillation a fraction, β , of the period T later than the first one. The second oscillation thus *lags behind* the first by a time βT , and so its displacement x_2 is given by

$$\begin{aligned} x_2 &= a \sin \omega(t - \beta T) \\ &= a \sin (\omega t - \varphi), \end{aligned} \quad (2)$$

where $\varphi = \omega \beta T = 2\pi \beta T/T = 2\pi \beta$. If the second oscillation *leads* the first by a time βT , the displacement is given by

$$x_2 = a \sin (\omega t + \varphi). \quad (3)$$

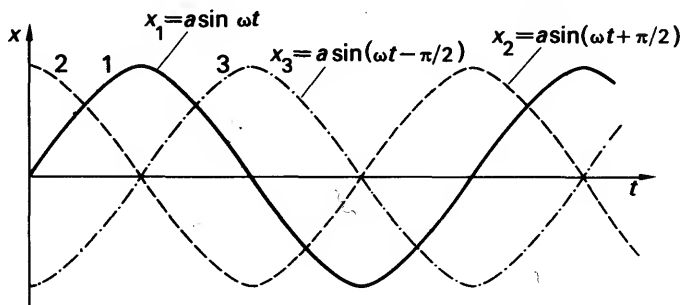


FIG. 25.3 Phase difference.

φ is known as the *phase angle* of the oscillation. It represents the *phase difference* between the oscillations $x_1 = a \sin \omega t$ and $x_2 = a \sin (\omega t - \varphi)$. Graphs of displacement v. time are shown in Fig. 25.3.

Curve 1 represents $x_1 = a \sin \omega t$. Curve 2 represents $x_2 = a \sin (\omega t + \pi/2)$, so that its phase lead is $\pi/2$; this is a lead of one quarter of a period. Curve 3 represents $x_3 = a \sin (\omega t - \pi/2)$ so that its phase lag is $\pi/2$; this is a lag of one quarter of a period on curve 1. If the phase difference is 2π , the oscillations are effectively in phase.

Note that if the phase difference is π , the displacement of one oscillation reaches a positive maximum value at the same instant as the other oscillation reaches a *negative* maximum value. The two oscillations are thus sometimes said to be 'antiphase'.

Damped Vibrations

In practice, the amplitude of vibration in simple harmonic motion does not remain constant but becomes progressively smaller. Such a vibration is said to be *damped*. The diminution of amplitude is due to loss of energy; for example, the amplitude of the bob of a simple pendulum diminishes slowly owing to the viscosity (friction) of the air. This is shown by curve 1 in Fig. 25.4.

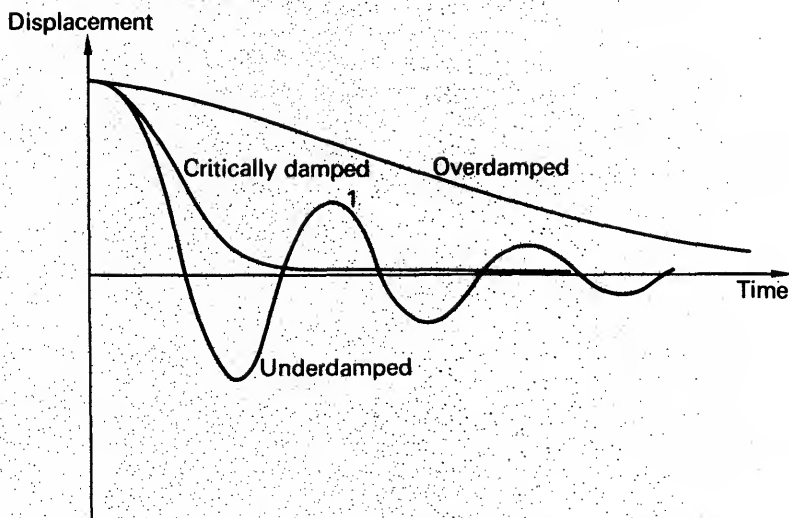


FIG. 25.4 Damped motion.

The general behaviour of mechanical systems subject to various amounts of damping may be conveniently investigated using a coil of a ballistic galvanometer (p. 919). If a resistor is connected to the terminals of a ballistic galvanometer when the coil is swinging, the induced emf due to the motion of the coil in the magnetic field of the galvanometer magnet causes a current to flow through the resistor. This current, by Lenz's Law (p. 896), opposes the motion of the coil and so causes damping. The smaller the value of the resistor, the greater is the degree of damping. The galvanometer coil is set swinging by discharging a capacitor through it. The time period, and the time taken for the amplitude to be reduced to a certain fraction of its original value, are then measured. The experiment can then be repeated using different values of resistor connected to the terminals.

It is found that as the damping is increased the time period increases and the oscillations die away more quickly. As the damping is increased further there is a value of resistance which is just sufficient to prevent the coil from vibrating past its rest position. This degree of damping, called the *critical damping*, reduces the motion to rest in the shortest possible time. If the resistance is lowered further, to increase the damping, no vibrations occur but the coil takes a longer time to settle down to its rest position. Graphs showing the displacement against time for 'underdamped', 'critically damped', and 'overdamped' motion are shown in Fig. 25.4.

When it is required to use a galvanometer as a current-measuring instrument, rather than ballistically to measure charge, it is generally critically damped. The return to zero is then as rapid as possible.

These results, obtained for the vibrations of a damped galvanometer coil, are quite general. All vibrating systems have a certain critical damping, which brings the motion to rest in the shortest possible time.

Forced Oscillations. Resonance

In order to keep a system, which has a degree of damping, in continuous oscillatory motion, some external periodic force must be used. The frequency of this force is called the *forcing frequency*. In order to see how systems respond to a forcing oscillation, we may use an electrical circuit comprising a coil L , capacitor C and resistor R , shown in Fig. 25.5 (see also p. 976).

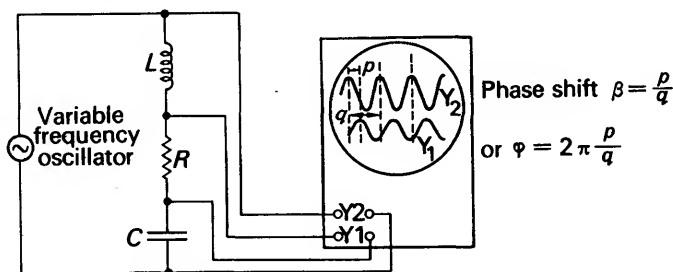


FIG. 25.5 Demonstration of oscillations.

The applied oscillating voltage is displayed on the Y_2 plates of a double-beam oscilloscope (p. 1011). The voltage across the resistor R is displayed on the Y_1 plates. Since the current I through the resistor is given by $I = V/R$, the voltage across R is a measure of the current through the circuit. The frequency of the oscillator is now set to a low value and the amplitude of the Y_1 display is recorded. The frequency is then increased slightly and the amplitude again measured. By taking many such readings, a graph can be drawn of the current through the circuit as the frequency is varied. A typical result is shown in Fig. 25.6 (i).

The phase difference, φ , between the Y_1 and Y_2 displays can be found by measuring the horizontal shift p between the traces, and the length q occupied by one complete waveform. φ is given by $(p/q) \times 2\pi$. A graph of the variation of phase difference between current and applied voltage can then be drawn. Fig. 25.6 (ii) shows a typical curve.

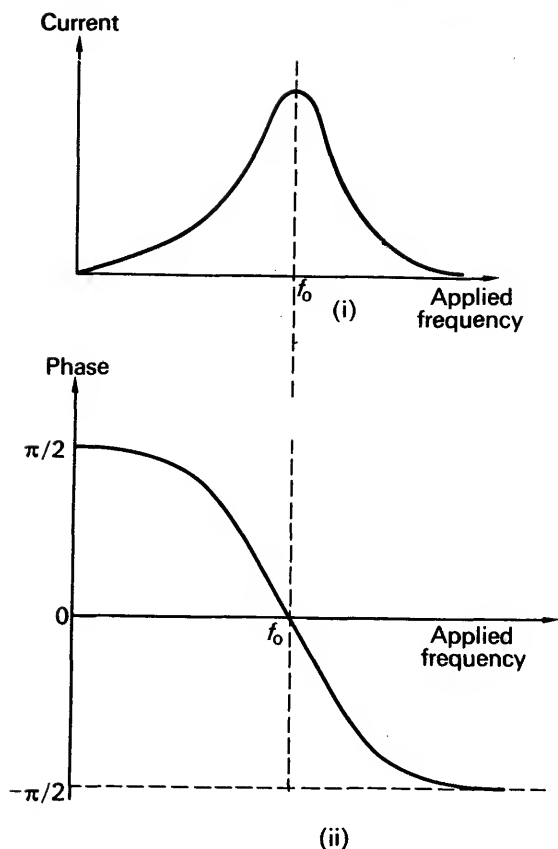


FIG. 25.6 Amplitude and phase in forced vibrations.

The following observations may be made:

1. The current is greatest at a certain frequency f_0 . This is the frequency of undamped oscillations of the system, when it is allowed to oscillate on its own. f_0 is called the *natural frequency* of the system. When the forcing frequency is equal to the natural frequency, *resonance* is said to occur. The largest current is then produced.

2. At resonance, the current and voltage are in phase. Well below resonance, the current leads the voltage by $\pi/2$; at very high frequencies the current lags by $\pi/2$. The behaviour of other resonant systems is similar.

3. The forced oscillations always have the same frequency as the forcing oscillations.

Examples of resonance occur in sound and in optics. These are discussed later. It should be noted that considerable energy is absorbed at the resonant frequency from the system supplying the external periodic force.

Waves and Wave-motion

A *wave* allows energy to be transferred from one point to another some distance away without any particles of the medium travelling between the two points. For example, if a small weight is suspended by a string, energy to move the weight may be obtained by repeatedly shaking the other end of the string up and down through a small distance. Waves, which carry energy, then travel along the string from the top to the bottom. Likewise, water waves may spread along the surface from one point A to another point B, where an object floating on the water will be disturbed by the wave. No particles of water at A actually travel to B in the process. The energy in the electromagnetic spectrum, comprising X-rays and light waves, for example, may be considered to be carried by electromagnetic waves from the radiating body to the absorber. Again, sound waves carry energy from the source to the ear by disturbance of the air (p. 585).

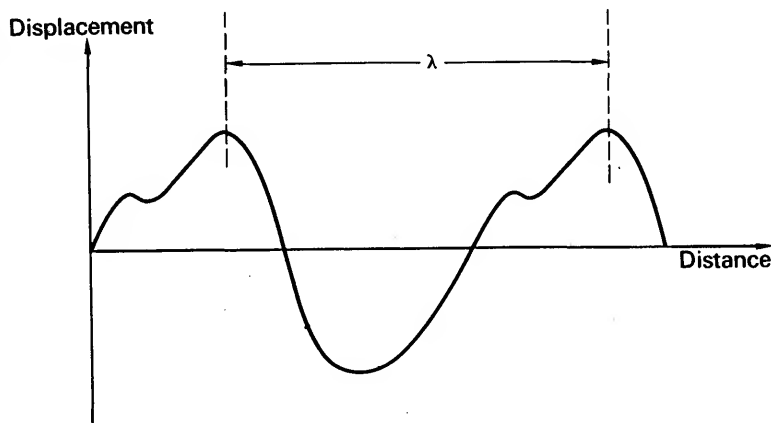


FIG. 25.7 Wave and wavelength.

If the source or origin of the wave oscillates with a frequency f , then each point in the medium concerned oscillates with the same frequency. A snapshot of the wave profile or waveform may appear as in Fig. 25.7 at a particular instant. The source repeats its motion f times per second, so a repeating *waveform* is observed spreading out from it. The distance between corresponding points in successive waveforms, such as two successive crests or two successive troughs, is called the *wavelength*, λ . Each time the source vibrates once, the waveform moves forward a distance λ . Thus in one second, when f vibrations occur, the wave moves forward a distance $f\lambda$. Hence the velocity v of the waves, which is the distance the profile moves in one second, is given by:

$$v = f\lambda.$$

This equation is true for all wave motion, whatever its origin, that is, it applies to sound waves, electromagnetic waves and mechanical waves.

Transverse Waves

A wave which is propagated by vibrations *perpendicular* to the direction of travel of the wave is called a *transverse* wave. Examples of transverse waves are waves on plucked strings and on water. Electro-magnetic waves, which include light waves, are transverse waves.

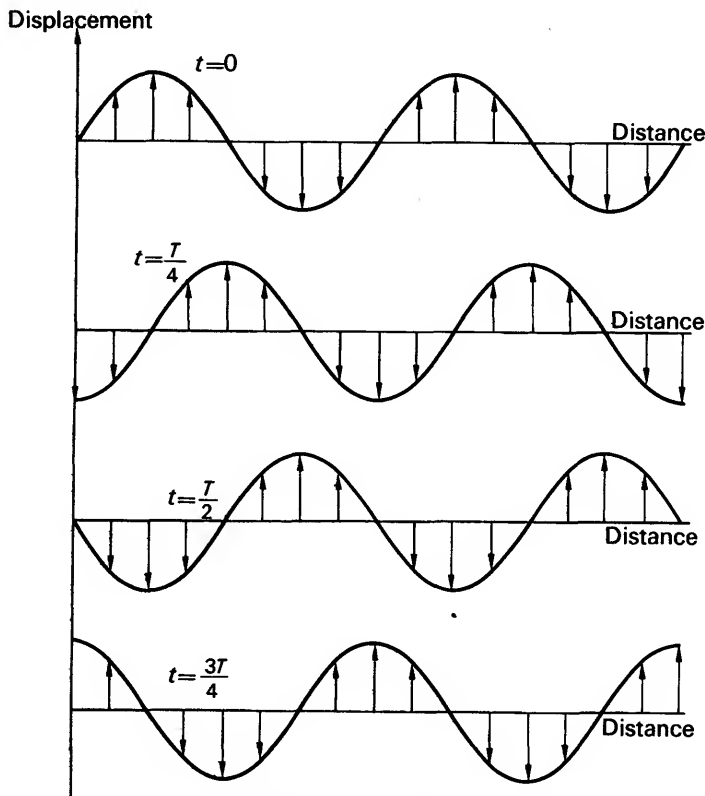


FIG. 25.8 Progressive transverse wave.

The propagation of a transverse wave is illustrated in Fig. 25.8. Each particle vibrates perpendicular to the direction of propagation with the same amplitude and frequency, and the wave is shown successively at $t = 0$, $T/4$, $T/2$, $3T/4$, in Fig. 25.8, where T is the period.

Longitudinal Waves

In contrast to a transverse wave, a *longitudinal* wave is one in which the vibrations occur in the *same* direction as the direction of travel of the wave. Fig. 25.9 illustrates the propagation of a longitudinal wave. The row of dots shows the actual positions of the particles whereas the *graph* shows the *displacement* of the particles from their equilibrium positions. The positions at time $t = 0$, $t = T/4$, $t = T/2$ and $t = 3T/4$

are shown. The diagram for $t = T$ is, of course, the same as $t = 0$. With displacements to R (right) and to L (left), note that:

(i) The displacements of the particles cause regions of high density (*compressions* C) and of low density (*rarefactions* R) to be formed.

(ii) These regions move along with the speed of the wave, as shown by the broken diagonal line.

(iii) Each particle vibrates about its mean position with the same amplitude and frequency.

(iv) The regions of greatest compression are one-quarter wavelength ahead of the greatest displacement in the direction of the wave. This result is important in understanding some processes involving sound waves.

The most common example of a longitudinal wave is a sound wave. This is propagated by alternate compressions and rarefactions of the air.

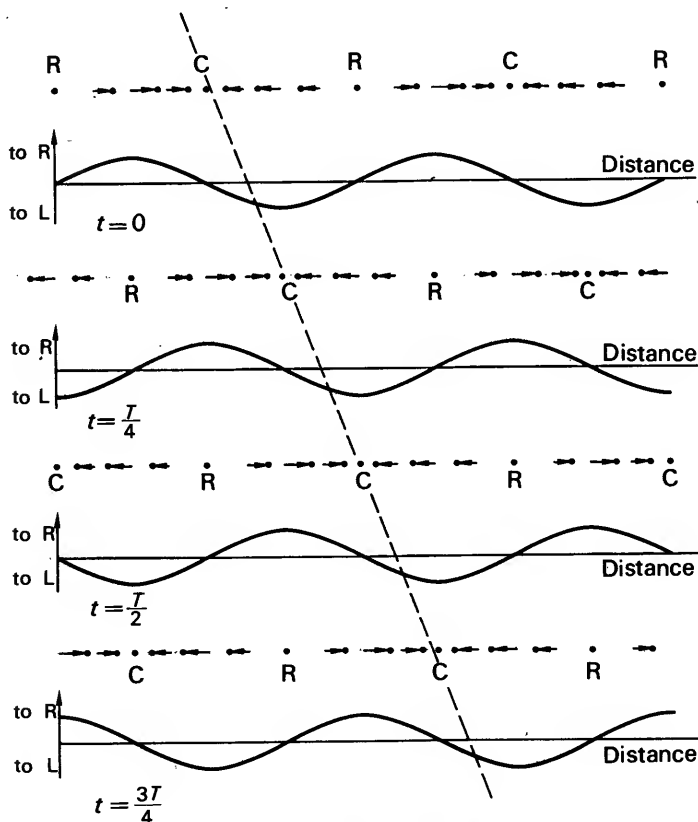


FIG. 25.9 Progressive longitudinal wave.

Progressive Waves

Both the transverse and longitudinal waves described above are *progressive*. This means that the wave profile moves along with the

speed of the wave. If a snapshot is taken of a progressive wave, it repeats at equal distances. The repeat distance is the *wavelength* λ . If one point is taken, and the profile is observed as it passes this point, then the profile is seen to repeat at equal intervals of time. The repeat time is the *period*, T .

The vibrations of the particles in a progressive wave are of the same amplitude and frequency. But *the phase of the vibrations changes for different points along the wave*. This can be seen by considering Figs. 25.8 and 25.9. The phase difference may be demonstrated by the following experiment.

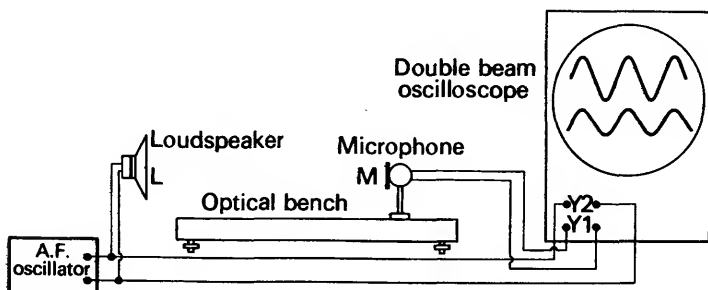


FIG. 25.10 Demonstration of phase in progressive wave.

An audio-frequency (af) oscillator is connected to the loudspeaker L and to the Y_2 plates of a double-beam oscilloscope, Fig. 25.10. A microphone M ,

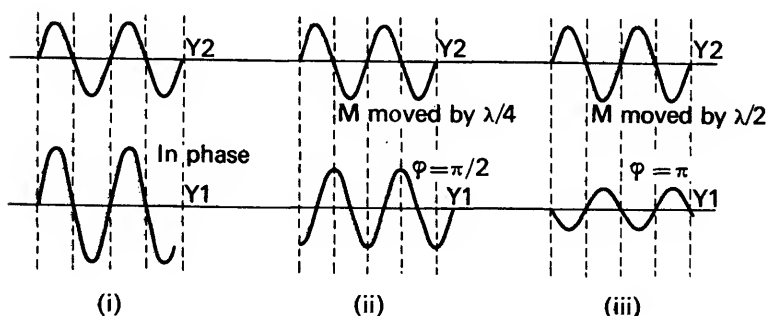


FIG. 25.11 Phase difference and wavelength.

mounted on an optical bench, is connected to the Y_1 plates. When M is moved away from or towards L , the two traces on the screen are as shown in Fig. 25.11 (i) at one position. This occurs when the distance LM is equal to a whole number of wavelengths, so that the signal received by M is in phase with that sent out by L . When M is now moved further away from L through a distance $\lambda/4$, where λ is the wavelength, the appearance on the screen changes to that shown in Fig. 25.11 (ii). The resultant phase change is $\pi/2$, so that the signal now arrives a quarter of a period later. When M is moved a distance $\lambda/2$ from its 'in-phase' position, the signal arrives half a period later, a phase change of π , Fig. 25.11 (iii).

Velocity of Sound in Free Air

The velocity of sound in free air can be found from this experiment. Firstly, a position of the microphone M is obtained when the two signals on the screen are in phase, as in Fig. 25.10. The reading of the position of M on the optical bench is then taken. M is now moved slowly until the phase of the two signals on the screen is seen to change through $\pi/2$ to π and then to be in phase again. The shift of M is then measured. It is equal to λ , the wavelength. From several measurements the average value of λ is found, and the velocity of sound is calculated from $v = f\lambda$, where f is the frequency obtained from the oscillator dial.

Progressive Wave Equation

An equation can be formed to represent generally the displacement y of a vibrating particle in a medium which a wave passes. Suppose the wave moves from left to right and that a particle at the origin O then vibrates according to the equation $y = a \sin \omega t$, where t is the time and $\omega = 2\pi f$ (p. 577).

At a particle P at a distance x from O to the right, the phase of the Displacement

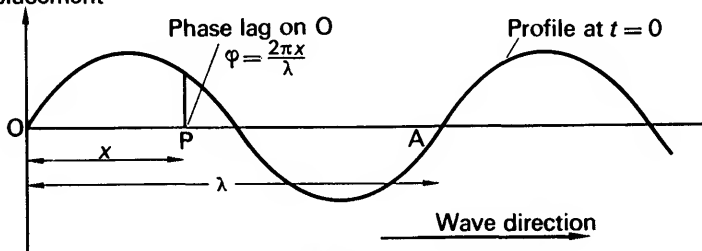


FIG. 25.12 Progressive wave equation.

vibration will be different from that at O, Fig. 25.12. A distance λ from O corresponds to a phase difference of 2π (p. 580). Thus the phase difference ϕ at P is given by $(x/\lambda) \times 2\pi$ or $2\pi x/\lambda$. Hence the displacement of any particle at a distance x from the origin is given by

$$y = a \sin (\omega t - \phi)$$

$$\text{or} \quad y = a \sin \left(\omega t - \frac{2\pi x}{\lambda} \right). \quad (4)$$

Since $\omega = 2\pi f = 2\pi v/\lambda$, where v is the velocity of the wave, this equation may be written:

$$y = a \sin \left(\frac{2\pi v t}{\lambda} - \frac{2\pi x}{\lambda} \right)$$

$$\text{or} \quad y = a \sin \frac{2\pi}{\lambda} (vt - x). \quad (5)$$

Also, since $\omega = 2\pi/T$, equation (4) may be written:

$$y = a \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda} \right). \quad (6)$$

Equations (5) or (6) represent a *plane-progressive wave*. The negative sign in the bracket indicates that, since the wave moves from left to right, the vibrations at points such as P to the right of O will lag on that at O. A wave travelling in the *opposite direction*, from right to left, arrives at P before O. Thus the vibration at P leads that at O. Consequently a wave travelling in the opposite direction is given by

$$y = a \sin 2\pi \left(\frac{t}{T} + \frac{x}{\lambda} \right), \quad (7)$$

that is, the sign in the bracket is now a plus sign.

As an illustration of calculating the constants of a wave, suppose a wave is represented by

$$y = a \sin \left(2000\pi t - \frac{\pi x}{17} \right),$$

where t is in seconds, y in cm. Then, comparing it with equation (5),

$$y = a \sin \frac{2\pi}{\lambda} (vt - x),$$

we have

$$\frac{2\pi v}{\lambda} = 2000\pi,$$

and

$$\frac{2\pi}{\lambda} = \frac{\pi}{17}.$$

$$\therefore \lambda = 2 \times 17 = 34 \text{ cm}$$

and

$$v = 1000\lambda = 1000 \times 34 = 34000 \text{ cm s}^{-1}$$

$$\therefore \text{frequency, } f, = \frac{v}{\lambda} = \frac{34000}{34} = 1000 \text{ Hz}$$

$$\therefore \text{period, } T, = \frac{1}{f} = \frac{1}{1000} \text{ s.}$$

If two layers of the wave are 180 cm apart, they are separated by $180/34$ wavelengths, or by $5\frac{10}{17}\lambda$. Their *phase difference* for a separation λ is 2π ; and hence, for a separation $10\lambda/34$, omitting 5λ from consideration, we have:

$$\text{phase difference} = \frac{10}{34} \times 2\pi = \frac{10\pi}{17} \text{ radians.}$$

Principle of Superposition

When two waves travel through a medium, their combined effect at any point can be found by the *Principle of Superposition*. This states that *the resultant displacement at any point is the sum of the separate displacements due to the two waves*.

The principle can be illustrated by means of a long stretched spring ('Slinky'). If wave pulses are produced at each end simultaneously, the two waves pass through the wire. Fig. 25.13(a) shows the stages which occur as the two pulses pass each other. In Fig. 25.13(a)(i), they are some distance apart and are approaching each other, and in Fig. 25.13(a)(ii)

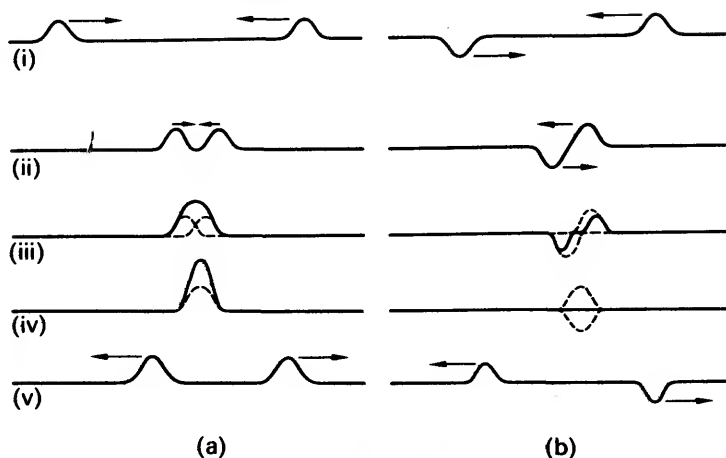


FIG. 25.13 Superposition of waves.

they are about to meet. In Fig. 25.13(a) (iii), the two pulses, each shown by broken lines, are partly overlapping. The resultant is the sum of the two curves. In Fig. 25.13(a) (iv), the two pulses exactly overlap and the greatest resultant is obtained. The last diagram shows the pulses receding from one another. The diagrams in Fig. 25.13(b) show the same sequence of events (i)–(v) but the pulses are equal and opposite. The Principle of Superposition is widely used in discussion of wave phenomena such as interference, as we shall see (p. 688).

Stationary Waves

We have already discussed progressive waves and their properties. Fig. 25.14 shows an apparatus which produces a different kind of wave (see also p. 662). If the weights on the scale-plan are suitably adjusted,

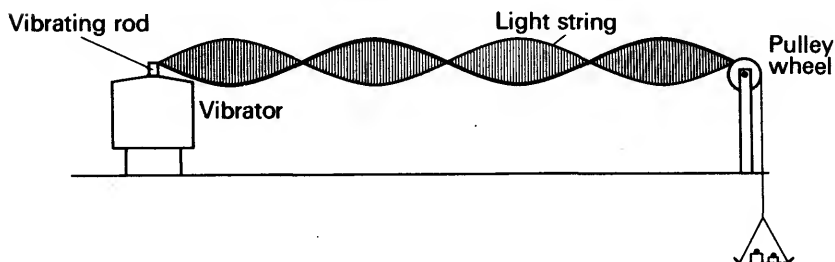


FIG. 25.14 Demonstration of stationary wave.

a number of *stationary vibrating loops* are seen on the string when one end is set vibrating. This time the wave-like profile on the string does *not* move along the medium, which is the string, and the wave is therefore called a *stationary* (or *standing*) wave.

The motion of the string when a stationary wave is produced can be seen by using a Xenon stroboscope (strobe). This instrument gives a

flashing light whose frequency can be varied. The apparatus is set up in a darkened room and illuminated with the strobe. When the frequency of the strobe is nearly equal to that of the string, the string can be seen moving up and down slowly. Its observed frequency is equal to the difference between the frequency of the strobe and that of the string. Progressive stages in the motion of the string can now be seen and studied, and these are illustrated in Fig. 25.15.

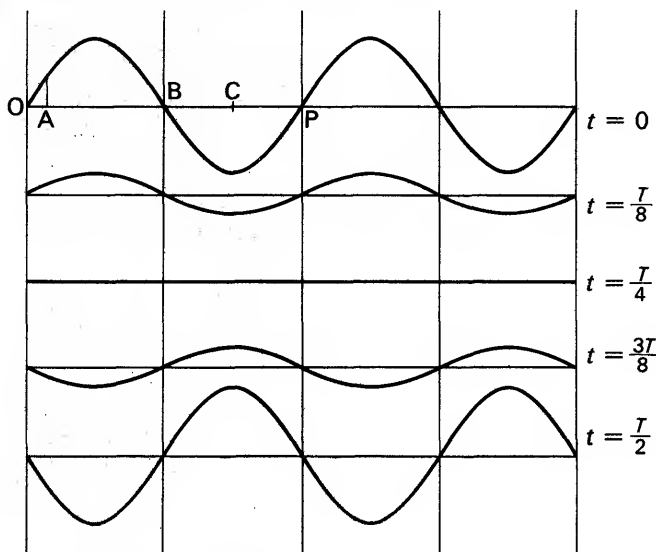


FIG. 25.15 Changes in motion of stationary wave.

The following points should be noted:

1. There are points such as B where the displacement is permanently zero. These points are called *nodes* of the stationary wave.

2. At points between successive nodes the vibrations *are in phase*. This property of the stationary wave is in sharp contrast to the progressive wave, where the phase of points near each other are all different (see p. 585). Thus when one point of a stationary wave is at its maximum displacement, *all* points are then at their maximum displacement. When a point (other than a node) has zero displacement, *all* points then have zero displacement.

3. Each point along the wave has a *different amplitude* of vibration from neighbouring points. Again this is different from the case of a progressive wave, where every point vibrates with the same amplitude. Points, e.g. C, which have the greatest amplitude are called *antinodes*.

4. The wavelength is equal to the distance OP, Fig. 25.15. Thus the wavelength λ , is twice the *distance between successive nodes or successive antinodes*. The distance between successive nodes or antinodes is $\lambda/2$; the distance between a node and a neighbouring antinode is $\lambda/4$.

Examples of stationary waves are discussed later in the book. The reader is referred to p. 641 for discussion of stationary waves in sound and to p. 985 for stationary electromagnetic waves.

Stationary Wave Equation

In deriving the wave equation of a progressive wave, we used the fact that the phase changes from point to point (p. 586). In the case of a stationary wave, we may find the equation of motion by considering the *amplitude* of vibration at each point because the amplitude varies while the phase remains constant.

Let ω be the angular frequency of the wave. The vibration of each particle may be represented by the equation

$$y = Y \sin \omega t, \quad . \quad . \quad . \quad . \quad (8)$$

where Y is the amplitude of the vibration at the point considered. Y varies along the wave with the distance x from some origin. If we suppose the origin to be at an antinode, then the origin will have the greatest amplitude, A , say. Now the wave repeats at every distance λ , and it can be seen that the amplitudes at different points vary sinusoidally with their particular distance x . An equation representing the changing amplitude Y along the wave is thus:

$$Y = A \cos \frac{2\pi x}{\lambda} = A \cos kx, \quad . \quad . \quad . \quad (9)$$

where $k = 2\pi/\lambda$. When $x = 0$, $Y = A$; when $x = \lambda$, $Y = A$. When $x = \lambda/2$, $Y = -A$. This equation hence correctly describes the variation in amplitude along the wave, Fig. 25.15. Hence the equation of motion of a stationary wave is, with (8),

$$y = A \cos kx \cdot \sin \omega t. \quad . \quad . \quad . \quad (10)$$

From (10), $y = 0$ at all times when $\cos kx = 0$. Thus $kx = \pi/2, 3\pi/2, 5\pi/2, \dots$, in this case. This gives values of x corresponding to $\lambda/4, 3\lambda/4, 5\lambda/4, \dots$. These points are *nodes* since the displacement at a node is always zero (p. 590). Thus equation (10) gives the correct distance, $\lambda/2$, between nodes.

A stationary wave can be considered as produced by the superposition of two progressive waves, of the same amplitude and frequency, travelling in opposite directions. This is shown mathematically on p. 643.

Wave Properties, Reflection and Refraction

Any wave motion can be *reflected*. The reflection of light and sound waves, for example, is discussed on pp. 677, 613 respectively.

Waves can also be *refracted*, that is, their direction changes when they enter a new medium. This is due to the change in velocity of the waves on entering a different medium. Refraction of light and sound, for example, is discussed later (pp. 679, 615).

Diffraction

Waves can also be 'diffracted'. *Diffraction* is the name given to the spreading of waves when they pass through apertures or around obstacles.

The general phenomenon of diffraction may be illustrated by using

water waves in a ripple tank, with which we assume the reader is familiar. Fig. 25.16(i) shows the effect of widening the aperture and Fig. 25.16(ii) the effect of shortening the wavelength and keeping the same width of opening. In certain circumstances in diffraction, reinforcement of the waves, or complete cancellation occurs in particular directions from the aperture, as shown in Fig. 25.16(i) and (ii). These patterns are called 'diffraction bands' (p. 701).

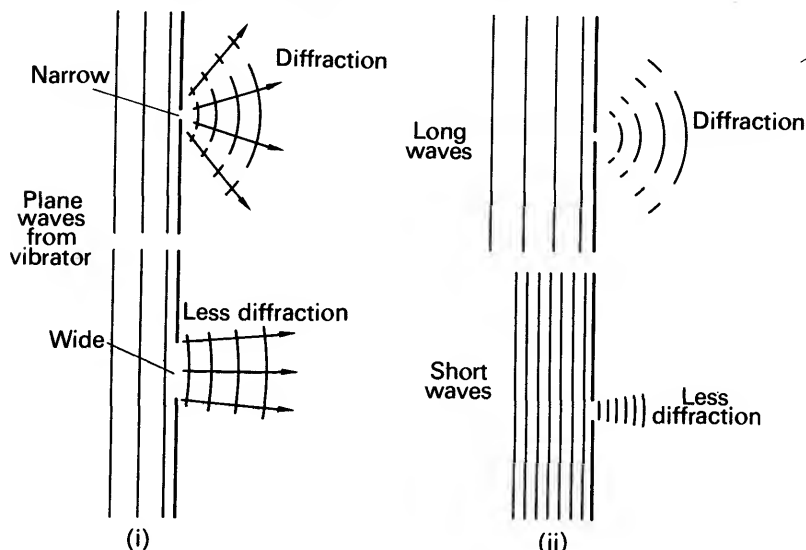


FIG. 25.16 Diffraction of waves.

Generally, the smaller the width of the aperture in relation to the wavelength, the greater is the spreading or diffraction of the waves. This explains why we cannot see round corners. The wavelength of *light waves* is about 6×10^{-5} cm (p. 690). This is so short that no appreciable diffraction is obtained around obstacles of normal size. With very small obstacles or narrow apertures, however, diffraction of light may be appreciable (see p. 707). *Sound waves* have long wavelengths, for example 50 cm, so that diffraction of sound waves occurs easily. For this reason, it is possible to hear round corners. *Electromagnetic waves* can be diffracted, as shown on p. 989.

Interference

When two or more waves of the same frequency overlap, the phenomenon of *interference* occurs. Interference is easily demonstrated in a ripple tank. Two sources, A and B, of the same frequency are used. These produce circular waves which spread out and overlap, and the pattern seen on the water surface is shown in Fig. 25.17.

The interference pattern can be explained from the Principle of Superposition (p. 588). If the oscillations of A and B are in phase, crests from A will arrive at the same time as crests from B at any point on the line RS. Hence by the Principle of Superposition there will be

reinforcement or a large wave along RS. Along XY, however, crests from A will arrive before corresponding crests from B. In fact, every point on XY is half a wavelength, λ , nearer to A than to B, so that crests from A arrive at the same time as troughs from B. Thus, by the Principle of Superposition, the resultant is *zero*. At every point along PQ there is a $3\lambda/2$ path difference from A compared to that from B, so that the resultant is also zero along PQ.

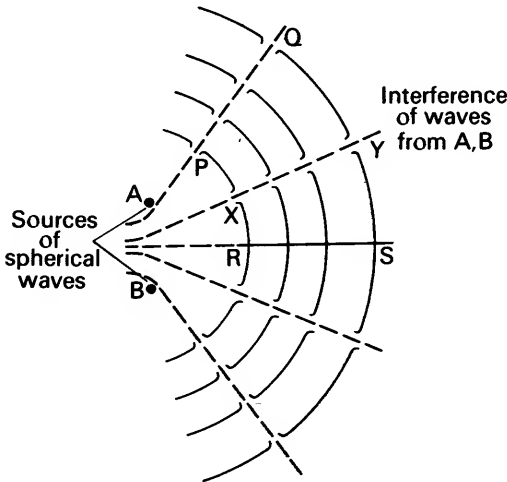


FIG. 25.17 Interference of waves.

Interference of *light* waves is discussed in detail on p. 687. An experiment to demonstrate the interference of *electromagnetic* waves (micro-waves) is given on p. 989. The interference of *sound* waves can be

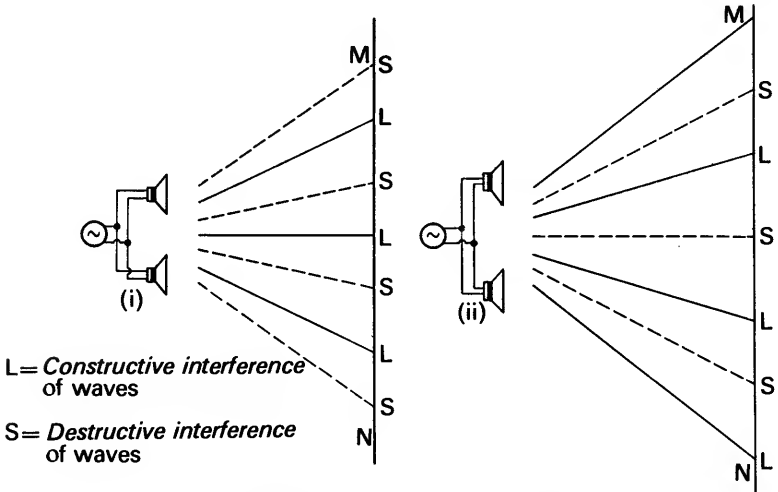


FIG. 25.18 Interference of sound waves.

demonstrated by connecting two loudspeakers in parallel to an audio-frequency oscillator, Fig. 25.18 (i). As the ear or microphone is moved along the line MN, alternate loud (L) and soft (S) sounds are heard according to whether the receiver of sound is on a line of reinforcement (constructive interference) or cancellation (destructive interference) of waves. Fig. 25.18 (i) indicates the positions of loud and soft sounds if the two speakers oscillate in phase. If the connections to *one* of the speakers is reversed, so that they oscillate out of phase, then the pattern is altered as shown in Fig. 25.18 (ii). The reader should try to account for this difference.

Velocity of Waves

We now list, for convenience, the velocity v of waves of various types, which are considered more fully in the appropriate sections of the book:

1. Transverse wave on string

$$V = \sqrt{\frac{T}{m}}, \quad . \quad . \quad . \quad . \quad (11)$$

where T is the tension and m is the mass per unit length.

2. Sound waves in gas

$$V = \sqrt{\frac{\gamma p}{\rho}}, \quad . \quad . \quad . \quad . \quad (12)$$

where p is the pressure, ρ is the density and γ is the ratio of the principal specific heat capacities of the gas.

3. Longitudinal waves in solid

$$V = \sqrt{\frac{E}{\rho}}, \quad . \quad . \quad . \quad . \quad (13)$$

where E is Young's modulus and ρ is the density.

4. Electromagnetic waves

$$V = \sqrt{\frac{1}{\mu\epsilon}}, \quad . \quad . \quad . \quad . \quad (14)$$

where μ is the permeability and ϵ is the permittivity of the medium.

SOUND RECEPTION, REPRODUCTION, RECORDING

The Ear

The eye can detect colour changes, which are due to the different frequencies of the light waves; it can also detect variations in brightness, which are due to the different amounts of light energy it receives. In the sphere of sound, the ear is as sensitive as the eye; it can detect notes of different pitch, which are due to the different frequencies of sound waves, and it can also detect loud and soft notes, which are due to different amounts of sound energy falling on the ear per second.

We are not concerned in this book with the complete physiology of the ear. Among other features, it consists of the *outer ear*, A, a canal C

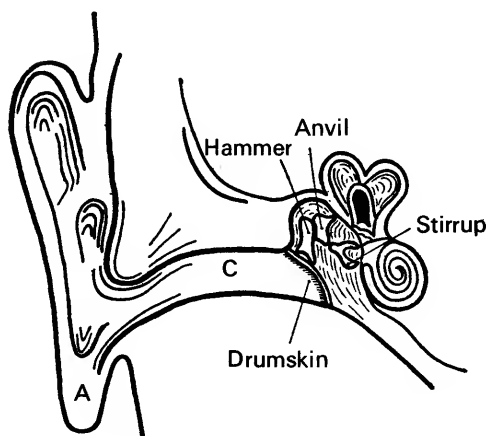


Fig. 25.19. The ear.

leading to the *drumskin*, and bones called the *ossicles*, Fig. 25.19. The ossicles consist of a bone called the 'hammer', fitting into another bone called the 'anvil', which is connected to the third bone called the 'stirrup'. When sound waves occur in the neighbourhood of the ear, they travel down the canal to the drumskin, which is also set into vibration. The part of the hammer in contact with the drumskin then vibrates, and strikes the anvil at the same rate. The motion is thus communicated to the stirrup, and from here it passes by a complicated mechanism to the auditory nerves, which set up the sensation of sound.

In 1843, OHM asserted that the ear perceives a simple harmonic vibration of the air as a simple or pure note. Although the waveforms of notes from instruments are far from being simple harmonic (see p. 610), the ear appears able to analyse a complicated waveform into the sum of a number of simple harmonic waves, which it then detects as separate notes.

Microphones

We now consider the principles of some *microphones*, instruments which convert sound energy to electrical energy. Details of microphones must be obtained from specialist works.

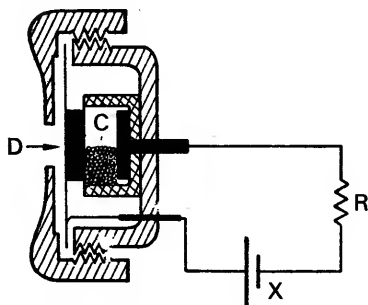


Fig. 25.20. Carbon microphone.

Carbon microphone, Fig. 25.20. This type of microphone is used in the hand set of a telephone. It contains carbon granules, C, whose electrical resistance decreases on compression and increases on release. Thus when sound waves are incident on the diaphragm D, a varying electric current of the same sound or audio frequency is produced along the telephone wires. The carbon microphone is a 'pressure' type since the magnitude of the current depends on the pressure changes in the air.

Ribbon microphone, Fig. 25.21. This is a sensitive microphone, with a uniform response over practically the whole of the audio-frequency (af) range from 40 to 15000 Hz. It has a corrugated aluminium ribbon R clamped between two pole-pieces N, S of a powerful magnet. When sound waves are incident on R, the ribbon vibrates perpendicular to the magnetic field. A varying induced emf of the same frequency is therefore obtained, as shown, and this is passed to an amplifier. The ribbon microphone is a 'velocity' type, since the movement of R depends on the velocity changes in the air particles near it.

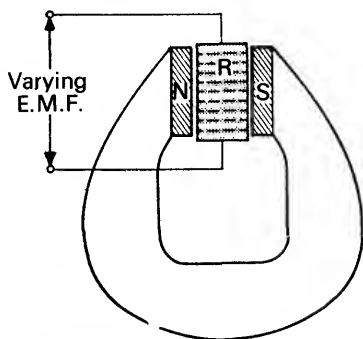


Fig. 25.21. Ribbon microphone.

Moving-coil microphone, Fig. 25.22. This is also a sensitive microphone, widely used. A coil X, made from aluminium tape so that it is very light, is situated in the radial field between the pole-pieces N, S of a magnet M. When sound waves are incident on a diaphragm Y

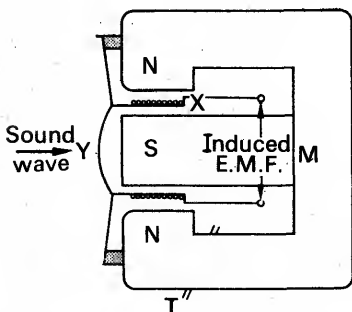


Fig. 25.22. Moving-coil microphone.

attached to X, the coil vibrates and 'cuts' magnetic flux. The varying induced emf obtained is passed to an amplifier. The moving-coil microphone is a pressure type. To avoid any effect of fluctuations in atmospheric pressure during use, the inside is kept at atmospheric pressure by means of a tube T.

The ribbon and moving-coil microphones are examples of *electrodynamical microphones*. They are widely used in the entertainment industry because their frequency response is extremely uniform. Unlike the carbon microphone, which has a relatively poor frequency response, no battery is used, and they are not subject to the background noise and 'hissing' obtained with a carbon microphone.

Loudspeaker. Telephone Earpiece

The *moving-coil loudspeaker* is used to reproduce sound energy from the electrical energy obtained with a microphone. It has a coil C or

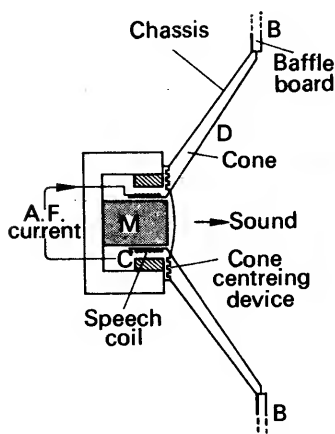


Fig. 25.23. Moving-coil loudspeaker.

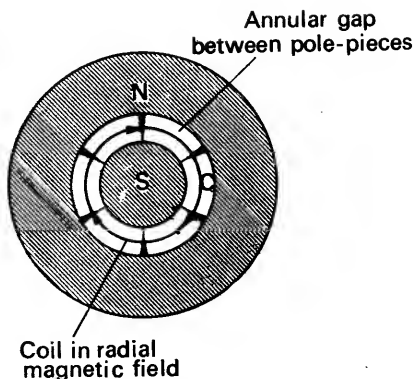


Fig. 25.24. Speech coil in magnetic field.

speech coil, wound on a cylindrical former, which is positioned symmetrically in the radial field of a pot magnet M, Fig. 25.23. A thin cardboard cone D is rigidly attached to the former and loosely connected to a large baffle-board B which surrounds it.

When C carries audio-frequency current, it vibrates at the same frequency in the direction of its axis. This is due to the force on a current-carrying conductor in a magnetic field, Fig. 25.24, whose direction is given by Fleming's left-hand rule. Since the surface area of the cone is large, the large mass of air in contact with it is disturbed and hence a loud sound is produced.

When the cone moves forward, a compression of air occurs in front of the cone and simultaneously a rarefaction behind it. The wave generated behind the cone is then 180° out of phase with that in front. If this wave reaches the front of the cone quickly, it will interfere appreciably with the wave there. Hence the intensity of the wave is diminished. This effect will be more noticeable at low frequency, or long wavelength, as the wave behind then has time to reach the front before the next vibration occurs. Generally, then, the sound would lack low note or bass intensity. The large baffle reduces this effect appreciably. It makes the path from the rear to the front so much longer that interference is negligible.

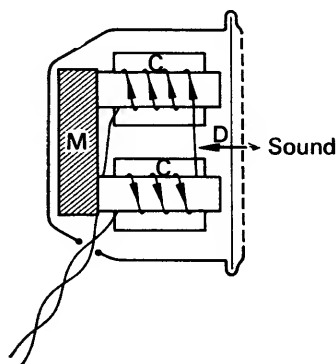


Fig. 25.25. Telephone earpiece.

Telephone earpiece, Fig. 25.25. This has speech coils C wound round soft-iron cores, which have a permanent magnet M between them. The varying speech current produces a corresponding varying attraction on the soft-iron diaphragm D, which thus generates sound waves of the

same frequency in the air. The sound is soft because the mass of air in contact with D is small. The permanent magnet is necessary to prevent distortion and to make the movement of D more sensitive to the varying attractive force.

Sound Recording and Reproduction

We now consider the principles of recording sound on tape and on film, and its reproduction. Details are beyond the scope of this book and must be obtained from manuals on the subject.

Tape recorder. Fig. 25.26 (i) illustrates the principle of tape recording in which flexible tape is used. It is coated with a fine uniform layer of a special form of ferric oxide which can be magnetised. The backing is a smooth plastic-base tape. When recording, the tape moves at a constant speed past the narrow gap between the poles of a ring of soft iron which has a coil round it. The coil carries the audio-frequency (af) current due to the sound recorded. On one half of a cycle, that part of the moving tape then in the gap is unmagnetised. On the other half of the same cycle, the next piece of tape in the gap is magnetised in the opposite direction, as shown. The rate at which pairs of such magnets is produced is equal to the frequency of the af current. The strength of the magnets is a measure of the magnetising current and hence of the intensity of the sound recorded.

In 'playback', the magnetised tape is now run at exactly the same speed past the same or another ring, the playback head, Fig. 25.26 (ii). As the small magnets pass the gap between the poles, the flux in the iron changes. An induced emf is thus obtained of the same frequency

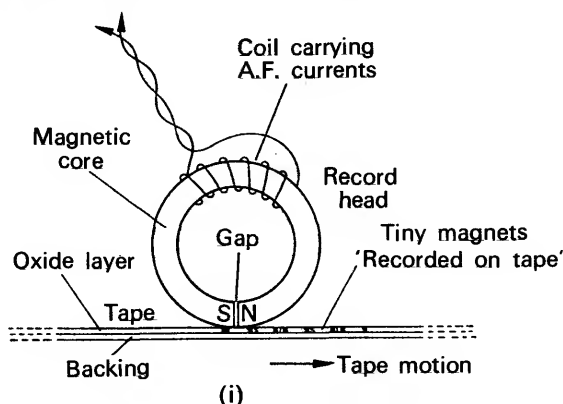


Fig. 25.26 (i). Tape recording.

and strength as that due to the original tape recording. This is amplified and passed to the loudspeaker, which reproduces the sound.

To obtain high-quality sound reproduction from tape, the output from a special high frequency *bias oscillator* is applied to the recording head in addition to the recording signal. This ensures that the magnets formed on the tape have strengths which are proportional to the recording signal. If the bias oscillator was not used, severe distortion due to

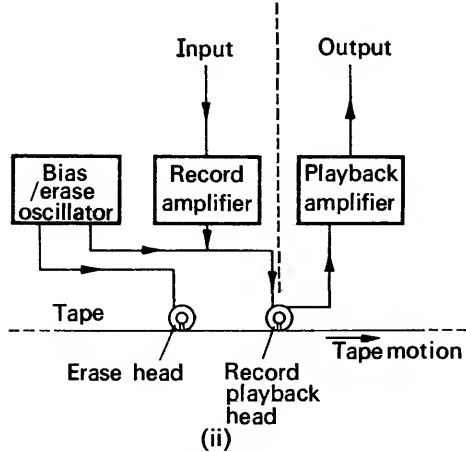


Fig. 25.26 (ii). Sound reproduction.

non-linearity would occur. The bias oscillator can also be used to *erase* the recording on the tape. This is done by applying its output to a coil in a special *erase head*. The erase head is similar to the record head but has a larger gap, so that the tape is in the magnetic field for a longer period. The bias signal takes the magnetic material on the tape through many thousands of hysteresis loops. These become progressively smaller until the magnetism disappears.

Sound Track

One method of recording sound on film, in the form of a *variable area* sound track, is illustrated in Fig. 25.27. A triangular aperture or mask T

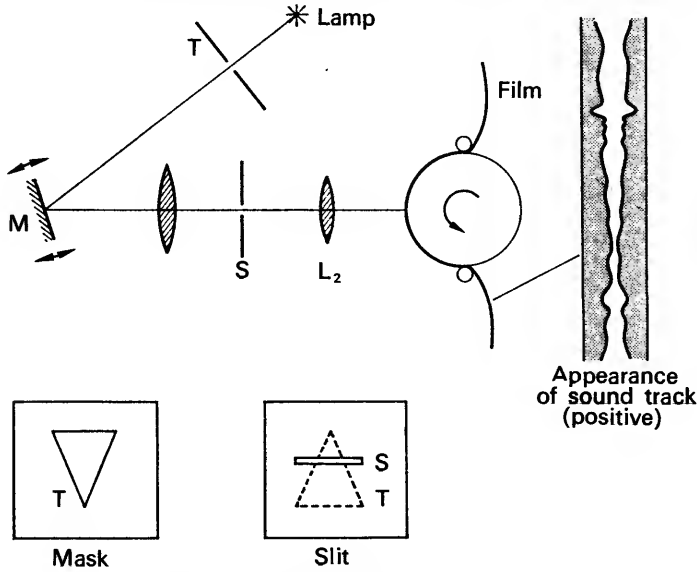


Fig. 25.27. Recording sound on film.

is brightly lit by a high-wattage lamp. After passing through T the light is reflected by a mirror M, and the rays are brought to a focus on to a slit S by a lens L_1 . By rotating the mirror slightly, as shown, the image T' of T, produced by the slit, can be moved up and down. This varies the length of slit illuminated. The light passing through the slit is collected by another lens system L_2 and focused on a strip of moving unexposed film. The mirror M is mounted on the moving system of a galvanometer, so that it moves at the same frequency as the audio-frequency currents passed through the galvanometer coil. The *area* of the film exposed thus varies as the audio-frequency signal. Hence when the film is developed, a permanent sound recording or *sound track* is obtained. A typical length of sound track is shown in Fig. 25.27.

Fig. 25.28 illustrates the principle of reproducing the sound. Light is focused on the sound track and passes through to a *photo-electric cell*. This contains a light-sensitive metal surface such as caesium, which then emits a number of electrons proportional to the light intensity. A current therefore flows in a resistor R. The sound track is coupled to the film, and as it moves, an audio-frequency current flows in R. The pd developed is amplified and passed to the loudspeaker.

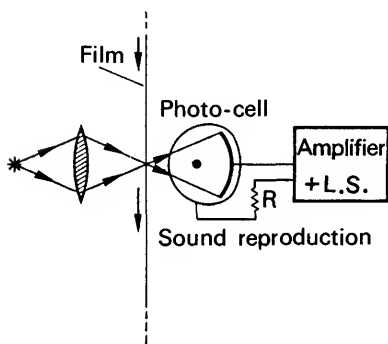


Fig. 25.28. Sound reproduction.

Frequency of Tuning-fork. Falling Plate Method

1. *Comparison method.* A tuning-fork is often used in experiments in sound to provide a note of known frequency. One method of measuring the frequency of a fork P is to compare it with the known frequency of another fork Q by a 'falling plate method'. One end of P is clamped in a vice; a light bristle B is attached to a prong, and rests lightly near the bottom of a smoked glass plate G, Fig. 25.29. The fork Q is similarly placed, and a bristle C, attached to one of its prongs, rests lightly on G. Both forks are sounded by drawing a bow across them, and the thread S suspending G is now burnt. The glass plate, usually in a groove, falls downward past the horizontally vibrating bristles, which then trace out two clear wavy 'tracks' BX, CY on G. Two horizontal lines, MN, are then drawn across BX, CY, and the number of complete waves between MN is counted on each trace. Suppose they are n_1 , n_2 respectively, and the frequencies of the corresponding forks are f_1 , f_2 . Then if t is the time taken by the plate to fall a distance MN,

$$f_1 = \text{number of cycles per second} = \frac{n_1}{t}$$

and $f_2 = n_2/t$. Hence $f_1/f_2 = n_1/n_2$, or

$$f_1 = \frac{n_1}{n_2} \times f_2.$$

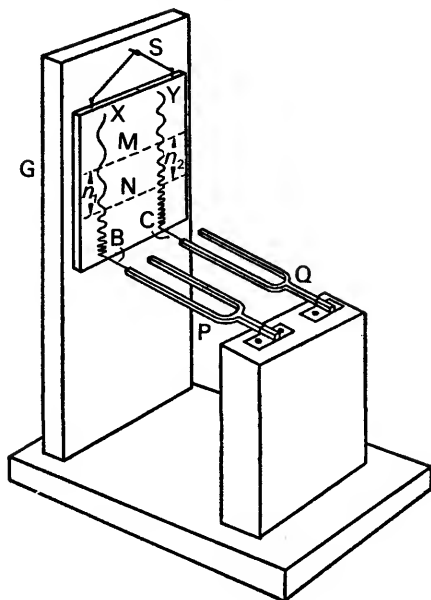


Fig. 25.29. Falling plate method.

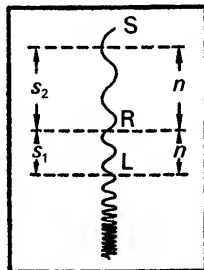
Thus, knowing n_1, n_2, f_2 , the unknown frequency f_1 can be calculated.

2. *Absolute method.* The unknown frequency f_1 of the tuning-fork P can also be calculated from its own trace. In this case the lengths s_1, s_2 corresponding to an *equal* number of consecutive waves are measured by a travelling microscope. Suppose there are n complete waves between LR, RS, Fig. 25.30. Then f_1 is given by

$$f_1 = n \sqrt{\frac{g}{s_2 - s_1}}, \quad (15)$$

where $g = 980$ when s_1, s_2 are in centimetres.

To prove this formula, for f_1 , let t be the time taken by the plate to fall a distance LR. Since the number of waves in LR is the same as in RS, the fork has also vibrated for a time t while the plate falls a distance LR. Thus if t is the time taken by the plate to fall a distance LR, $2t$ is the time it takes to fall a distance LS. Suppose u is the velocity of the plate at the instant when the line L on it reaches the vibrating bristle on the fork.


 Fig. 25.30.
(Not to Scale.)

Then

$$s_1 = ut + \frac{1}{2}gt^2 \quad (i)$$

from the dynamics equation $s = ut + \frac{1}{2}ft^2$, since the acceleration, f , of the plate $= g$, the acceleration due to gravity. As the time taken by the plate to fall a distance LS, or $(s_1 + s_2)$, is $2t$, we have

$$s_1 + s_2 = u \cdot 2t + \frac{1}{2}g(2t)^2. \quad (ii)$$

Multiplying equation (i) by 2, and subtracting from (ii) to eliminate u ,

$$\begin{aligned} s_2 - s_1 &= gt^2, \\ \therefore t &= \sqrt{\frac{s_2 - s_1}{g}} \\ \therefore f_1 &= \frac{n}{t} = n\sqrt{\frac{g}{s_2 - s_1}} \end{aligned} \quad (16)$$

as given above in equation (15).

The falling-plate method gives only an approximate value of the frequency, as (a) it is difficult to determine an *exact* number of waves; (b) the attachment of a bristle to the tuning-fork prong lowers its frequency slightly (see p. 622); (c) there is friction between the style and the plate.

The Stroboscope

A *stroboscope* is an arrangement which can make a rotating object appear at rest when it is viewed, and thus enables a spinning wheel, for example, to be studied at leisure. The stroboscopic method can be used to determine the frequency of a tuning-fork, which is electrically maintained for the purpose.

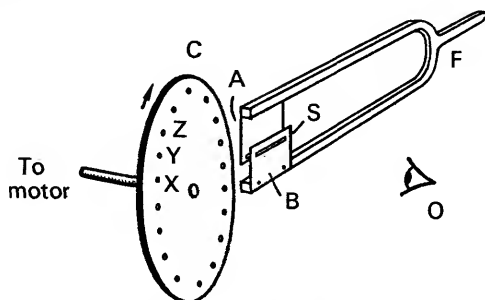


Fig. 25.31. Stroboscope method.

Two light metal plates, A, B, each with a slit S in them, are attached to prongs of the tuning-fork F so that the slits overlap each other when the fork is not sounded, Fig. 25.31. Behind the slit is a vertical circular white card C with black dots spaced at equal distances round the circumference, and the dots on the card can be seen through S. The tuning-fork is set into vibration, and the card is rotated by a motor about a horizontal axis through its centre with increasing speed. At first an observer O, viewing the dots through S, sees them moving round in an opposite direction to that in which the card is rotated. This is because the intermittent glimpses of the card through S occur quicker than the time taken for one dot to reach the place of the dot in front of it, with the result that the dots appear to be moving slowly back. As the speed of the card is increased further, a stage is reached when the dots appear perfectly stationary.

Through the slit S, glimpses of the dots are seen twice in every cycle

of vibration of the tuning-fork. When the dots first appear stationary, a particular dot such as X moves to a neighbouring dot position Y at one glimpse of the wheel, then to the next dot position Z at the next glimpse, and so on. At the end of one second, $2f$ glimpses have occurred, where f is the frequency of the fork in Hz. If the wheel has m dots and is now rotating at n rev s^{-1} , there have been $m \times n$ dot successive movements in one second.

$$\therefore 2f = mn, \text{ or } f = \frac{mn}{2}. \quad (17)$$

At twice the speed of revolution, $2n$ rev s^{-1} , the dots are again seen stationary. But this time only half their number is seen, as a particular dot moves through two dot places between successive glimpses. The dots may again be seen stationary at $3n, 4n, \dots$ rev s^{-1} of the wheel if the stress on the wheel at higher speeds is below a dangerous level.

As an illustration, suppose a wheel has 40 equally spaced dots and is viewed stroboscopically by a fork of 300 Hz. If the dots are seen stationary at the lowest angular speed, the number of revs per sec, n , is given by

$$600 = n \times 40, \text{ or } n = 15 \text{ rev } s^{-1}.$$

A neon lamp, providing intermittent flashes of light at a rate which can be varied by an electrical circuit, is used as a stroboscope in industry to adjust critically the speed of rotating wheels or machinery, which then appear stationary. The wear and tear with time of the moving parts of watches have been photographed with the aid of a stroboscope.

Relative movement of dots. It is sometimes impossible to keep the speed of rotation of the wheel in Fig. 25.31 constant, so that the dots appear to move slowly forward or back. Suppose, for example, that 2 dots per second cross the line of view in a backward direction in the case of the fork and wheel just discussed. Instead of 40×15 or 600 successive dot movements at 15 revs per second, when the wheel appears stationary, there are now $600 - 2$ or 598 successive dot movements. Since there are 40 dots round the wheel,

$$\therefore \text{new rate of revolution of wheel} = \frac{598}{40} = 14.95 \text{ rev } s^{-1}.$$

Suppose that the 40 dots appear stationary again at a wheel rotation of 15 revs per sec when viewed stroboscopically with the fork of frequency 300 Hz, and the fork is now loaded with a small piece of plasticine. The fork frequency is then lowered to a value f' . In this case the time interval between successive glimpses is longer than before, so that the dots appear to move forward. Suppose the movement is 3 dots per 10 seconds across the field of view. The number of successive glimpses of the wheel is $10 \times 2f'$ in 10 seconds. In this time the number of successive dot movements is $10 \times 40 \times 15 - 3$, or 5997.

$$\therefore 20f' = 5997$$

$$\therefore f' = 299.85.$$

Thus the frequency of the fork is lowered by 0.15 Hz.

EXERCISES 25

1. A tuning-fork is considered to produce a 'pure' note. (i) Write down an equation which represents the vibration of the prongs. (ii) Explain how an *exchange of energy* occurs during the motion of the prongs.

2. State, with reasons, whether the following waves are *longitudinal* (L) or *transverse* (T): *sound, waves on plucked string, water waves, light waves*. Draw sketches to illustrate your answer.

3. If the velocity of sound in air is 340 metres per second, calculate (i) the wavelength in cm when the frequency is 256 Hz, (ii) the frequency when the wavelength is 85 cm.

4. A plane-progressive wave is represented by the equation

$$y = 0.1 \sin (200\pi t - 20\pi x/17),$$

where y is the displacement in millimetres, t is in seconds and x is the distance from a fixed origin O in metres (m).

Find (i) the frequency of the wave, (ii) its wavelength, (iii) its speed, (iv) the phase difference in radians between a point 0.25 m from O and a point 1.10 m from O, (v) the equation of a wave with double the amplitude and double the frequency but travelling exactly in the opposite direction.

5. Describe how a *sound wave* passes through air, using graphs which illustrate and compare the variation of (i) the *displacement* of the air particles, (ii) the *pressure changes*, while the wave travels.

6. In the *falling-plate* experiment of measuring the frequency of a tuning-fork, the successive lengths occupied by 25 complete waves were 9.85 and 14.1 cm. When the experiment was repeated, the successive lengths occupied by 25 complete waves were 9.0 and 13.0 cm. Calculate the frequency of the fork for each experiment, deriving any formula you employ.

7. Describe an absolute method for determining the frequency of a tuning-fork.

Two forks A and B vibrate in unison but when two slits are fixed to the prongs of A , so that they are in line when the prongs are at rest, 9 beats in 10 sec. are heard when the forks are sounded together. A is then made to vibrate in front of a stroboscopic disc, on which are marked 50 equally spaced radial lines. The disc is viewed through the slits and the lines appear at rest when the disc rotates at 25 rev s^{-1} . What is the frequency of B ?

Describe and explain what would be seen if the speed of rotation of the disc were slightly decreased. (L.)

8. Describe the nature of the disturbance set up in air by a vibrating tuning-fork and show how the disturbance can be represented by a sine curve. Indicate on the curve the points of (a) maximum particle velocity, (b) maximum pressure.

What characteristics of the vibration determine the pitch, intensity, and quality respectively of the note? (N.)

9. Distinguish between *longitudinal* and *transverse* wave motions, giving examples of each type. Find a relationship between the frequency, wavelength and velocity of propagation of a wave motion.

Describe experiments to investigate quantitatively for sound waves the phenomena of (a) reflection, (b) refraction, (c) interference. (C.)

10. State and explain the differences between progressive and stationary waves.

A progressive and a stationary simple harmonic wave each have the same frequency of 250 Hz and the same velocity of 30 m s^{-1} . Calculate (i) the phase difference between two vibrating points on the progressive wave which are 10 cm apart, (ii) the equation of motion of the progressive wave if its amplitude is 0.03 metre, (iii) the distance between nodes in the stationary wave, (iv) the equation of motion of the stationary wave if its amplitude is 0.01 metre.

11. (i) Describe a stroboscope and an experiment to demonstrate one of its uses. Explain the calculation involved. (ii) Explain how sound is recorded and reproduced in a tape recorder. (L.)

12. Describe a *ribbon* and a *moving-coil* microphone, and explain with the aid of a diagram how each functions.

Draw a labelled diagram of the principal features of a *moving-coil loud-speaker*. Explain the purpose of the *baffle-board*.

13. Give a brief account of the principle of the stroboscope.

Describe how such a device may be used to determine the frequency of a tuning-fork. (You may, if necessary, suppose that two forms of equal frequency are available.)

Give a short account of any use made outside the laboratory of the stroboscope principle.

When viewed stroboscopically at a frequency of 300 vibrations second^{-1} a circular disc with 40 equally spaced dots appears to have a backwards rotation such that two dots cross the viewing line each second. What is the least rate of rotation of the disc? (N.)

14. Explain the terms *damped oscillation*, *forced oscillation* and *resonance*. Give one example of each.

Describe an experiment to illustrate the behaviour of a simple pendulum (or pendulums) undergoing forced oscillation. Indicate qualitatively the results you would expect to observe.

What factors determine (a) the period of free oscillations of a mechanical system, and (b) the amplitude of a system undergoing forced oscillation? (O. & C.)

chapter twenty-six

Characteristics, properties, and velocity of sound waves

CHARACTERISTICS OF NOTES

NOTES may be similar to or different from each other in three respects: (i) *pitch*, (ii) *loudness*, (iii) *quality*; so that if each of these three quantities of a particular note is known, the note is completely defined or "characterised".

Pitch

Pitch is analogous to colour in light, which is characterised by the wavelength, or by the frequency, of the electromagnetic vibrations (p. 690). Similarly, the pitch of a note depends only on the frequency of the vibrations, and a high frequency gives rise to a high-pitched note. A low frequency produces a low-pitched note. Thus the high-pitched whistle of a boy may have a frequency of several thousand Hz, whereas the low-pitched hum due to A.C. mains frequency when first switched on may be a hundred Hz. The range of sound frequencies is about 15 to 20000 Hz and depends on the observer.

Musical Intervals

If a note of frequency 300 Hz, and then a note of 600 Hz, are sounded by a siren, the pitch of the higher note is recognised to be an upper octave of the lower note. A note of frequency 1000 Hz is recognised to be an upper octave of a note of frequency 500 Hz, and thus the *musical interval* between two notes is an upper octave if the ratio of their frequencies is 2:1. It can be shown that the musical interval between two notes depends on the *ratio* of their frequencies, and not on the actual frequencies. The table below shows the various

Note	C	D	E	F	G	A	B	c
	doh	ray	me	fah	soh	lah	te	doh
Natural (Diatonic) Scale								
Frequency	256	288	320	341	384	427	480	512
Intervals between notes ..	9/8		10/9	16/15	9/8	10/9	9/8	16/15
Intervals above C ..	1.000	1.125	1.250	1.333	1.500	1.667	1.875	2.000
Equal Temperament Scale								
intervals above C ..	1.000	1.122	1.260	1.335	1.498	1.682	1.888	2.000

Note 1.—There are 12 semitones to the octave in the scale of equal temperament; each semitone has a frequency ratio of $2^{1/12}$.

Note 2.—The frequency of C is 256 on the scale of Helmholtz above; in music it is 261.2.

musical intervals and the corresponding ratio of the frequencies of the notes.

Intensity and Loudness

The *intensity* of a sound at a place is defined as the energy per second flowing through one square metre held normally at that place to the direction along which the sound travels. As we go farther away from a source of sound the intensity diminishes, since the intensity decreases as the square of the distance from the source (see also p. 565).

Suppose the displacement y of a vibrating layer of air is given by $y = a \sin \omega t$, where $\omega = 2\pi/T$ and a is the amplitude of vibration, see equation (1), p. 577. The velocity, v , of the layer is given by

$$v = \frac{dy}{dt} = \omega a \cos \omega t,$$

and hence the kinetic energy, W , is given by

$$W = \frac{1}{2} m v^2 = \frac{1}{2} m \omega^2 a^2 \cos^2 \omega t \quad . \quad . \quad . \quad (i)$$

where m is the mass of the layer. The layer also has potential energy as it vibrates. Its total energy, W_0 , which is constant, is therefore equal to the maximum value of the kinetic energy. From (i), it follows that

$$W_0 = \frac{1}{2} m \omega^2 a^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (ii)$$

In 1 second, the air is disturbed by the wave over a distance V cm., where V is the velocity of sound in m s^{-1} ; and if the area of cross-section of the air is 1 m^2 , the volume of air disturbed is $V \text{ m}^3$. The mass of air disturbed per second is thus $V\rho$ kg, where ρ is the density of air in kg m^{-3} , and hence, from (ii),

$$W = \frac{1}{2} V \rho \omega^2 a^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (iii)$$

It therefore follows that *the intensity of a sound due to a wave of given frequency is proportional to the square of its amplitude of vibration.*

It can be seen from (ii) that the greater the mass m of air in vibration, the greater is the intensity of the sound obtained. For this reason the sound set up by the vibration of the diaphragm of a telephone earpiece cannot be heard except with the ear close to the earpiece. On the other hand, the cone of a loudspeaker has a large surface area, and thus disturbs a large mass of air when it vibrates, giving rise to a sound of much larger intensity than the vibrating diaphragm of the telephone earpiece. It is difficult to hear a vibrating tuning-fork a small distance away from it because its prongs set such a small mass of air vibrating. If the fork is placed with its end on a table, however, a much louder sound is obtained, which is due to the large mass of air vibrating by contact with the table.

Loudness is a sensation, and hence, unlike intensity, it is difficult to measure because it depends on the individual observer. Normally, the greater the intensity, the greater is the loudness of the sound (see p. 607).

The Decibel

We are already familiar with the fact that when the frequency of a note is doubled its pitch rises by an octave. Thus the increase in pitch

sounds the same to the ear when the frequency increases from 100 to 200 Hz as from 500 to 1000 Hz. Similarly, it is found that increases in loudness depend on the *ratio* of the intensities, and not on the absolute differences in intensity.

If the power of a source of sound increases from 0.1 watt to 0.2 watt, and then from 0.2 watt to 0.4 watt, the loudness of the source to the ear increases in equal steps. The equality is thus dependent on the equality of the *ratio* of the powers, not their difference, and in commercial practice the increase in loudness is calculated by taking the *logarithm of the ratio of the powers to the base 10*, which is $\log_{10} 2$, or 0.3, in this case.

Relative intensities or powers are expressed in *bels*, after Graham Bell, the inventor of the telephone. If the power of a source of sound changes from P_1 to P_2 , then

$$\text{number of bels} = \log_{10} \left(\frac{P_2}{P_1} \right).$$

In practice the bel is too large a unit, and the **decibel (db)** is therefore adopted. This is defined as one-tenth of a bel, and hence in the above case

$$\text{number of decibels} = 10 \log_{10} \left(\frac{P_2}{P_1} \right).$$

The minimum change of power which the ear is able to detect is about 1 db, which corresponds to an increase in power of about 25 per cent.

Calculation of Decibels

Suppose the power of a sound from a loudspeaker of a radio receiver is 50 milliwatts, and the volume control is turned so that the power increases to 500 milliwatts. The increase in power is then given by

$$10 \log_{10} \left(\frac{P_2}{P_1} \right) = 10 \times \log \frac{500}{50} = 10 \text{ db.}$$

If the volume control is turned so that the power increases to 1000 milliwatts, the increase in power compared with the original sound

$$= 10 \log_{10} \left(\frac{1000}{50} \right) = 10 \log_{10} 20 = 13 \text{ db.}$$

If the volume control is turned down so that the power decreases from 1000 to 200 milliwatts, the change in power

$$= 10 \log_{10} \left(\frac{200}{1000} \right) = 10 \log_{10} 2 - 10 \log_{10} 10 = -7 \text{ db.}$$

The minus indicates a decrease in power. Besides its use in acoustics the decibel is used by radio and electrical engineers in dealing with changes in electrical power.

Intensity Levels. Threshold of Hearing

Since the intensity of sound is defined as the energy per second crossing 1 metre² normal to the direction of the sound, the unit of intensity is "watt metre⁻²", symbol "W m⁻²". The *intensity level* of a source is its

intensity relative to some agreed 'zero' intensity level. If the latter has an intensity of P_0 watt metre⁻², a sound of intensity P watt metre⁻² has an intensity level defined as:

$$10 \log_{10} \left(\frac{P}{P_0} \right) \text{ db.}$$

The lowest audible sound at a frequency of 1000 Hz, which is called the *threshold of hearing*, corresponds to an intensity P_0 of 10^{-12} watt m⁻² or 10^{-10} microwatt cm⁻². This is chosen as the 'zero' of sound intensity level. An intensity level of a low sound of + 60 db is 60 decibels or 6 bels higher than 10^{-12} watt metre⁻². The intensity is thus 10^6 times as great, and is therefore equal to $10^6 \times 10^{-12}$ or 10^{-6} W m⁻².

Calculation of intensity level. The difference in intensity levels of two sounds of intensities P_1, P_2 watt metre⁻² respectively is $10 \log_{10}(P_2/P_1)$ db. Thus the difference in intensity levels of a sound of intensity 8×10^{-5} W m⁻² due to a person talking, and one of an intensity 10^{-1} W m⁻² due to an orchestra playing,

$$= 10 \log_{10} \left(\frac{8 \times 10^{-5}}{10^{-1}} \right) = -40 + 10 \log_{10} 8 = -31 \text{ db.}$$

The negative sign indicates a decrease in intensity level. Similarly, if two intensity levels differ by 20 decibels, the ratio P'/P of the two intensities is given by

$$10 \log_{10} \left(\frac{P'}{P} \right) = 20,$$

or

$$\frac{P'}{P} = 10^2 = 100.$$

A source of sound such as a small loudspeaker produces a sound intensity round it proportional to $1/d^2$, where d is the distance from the loudspeaker (p. 607). Suppose the intensity level is 10 db at a distance of 20 m from the speaker. At a distance of 40 m the intensity will be four times less than at 20 m, a reduction of $10 \log_{10} 4$ db or about 6 db.

\therefore intensity level here = 10 db - 6 db = 4 db.

At a point 10 m from the speaker the intensity will be four times greater than at 20 m. The intensity level here is thus 10 + 6 or 16 db.

If the electrical power supplied to the loudspeaker is doubled, the sound intensity at each point is doubled. Thus if the original intensity level was 16 db, the new intensity level is higher by $10 \log_{10} 2$ db or about 3 db. The new intensity level is hence 19 db.

Loudness. The Phon

The loudness of a sound is a sensation, and thus depends on the observer, whereas power, or intensity, of a sound is independent of the observer. Observations show that sounds which appear equally loud to a person have different intensities or powers, depending on the frequency, f , of the sound. The curves a, b, c represent respectively three values of *equal loudness*, and hence the intensity at X, when the frequency is 1000 Hz, is less than the intensity at Y, when the frequency is 500, although the loudness is the same, Fig. 26.1.

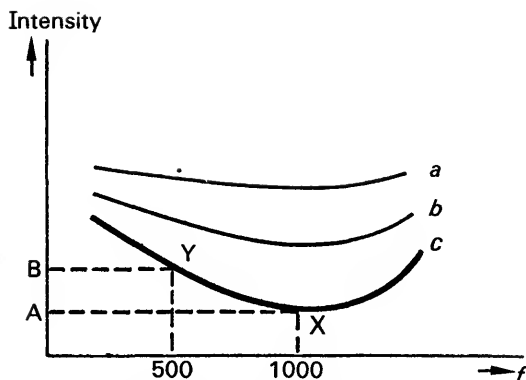


FIG. 26.1. Curves of equal loudness.

In order to measure loudness, therefore, scientists have adopted a “standard” source having a frequency of 1000 Hz, with which all other sounds are compared. The source *H* whose loudness is required is placed near the standard source, and the latter is then altered until the loudness is the same as *H*. The intensity or power level of the standard source is then measured, and if this is n decibels above the threshold value (10^{-10} microwatt per sq cm, p. 609) the loudness is said to be n phons. The phon, introduced in 1936, is thus a unit of loudness, whereas the decibel is a unit of intensity or power. *Noise meters*, containing a microphone, amplifier, and meter, are used to measure loudness, and are calibrated directly in phons. The “threshold of feeling”, when sound produces a painful sensation to the ear, corresponds to a loudness of about 120 phons.

Quality or Timbre

If the same note is sounded on the violin and then on the piano, an untrained listener can tell which instrument is being used, without seeing it. We say that the *quality* or *timbre* of the note is different in each case.

The waveform of a note is never simple harmonic in practice; the nearest approach is that obtained by sounding a tuning-fork, which thus produces what may be called a “pure” note, Fig. 26.2 (i). If the same note is played on a violin and piano respectively, the waveforms produced might be represented by Fig. 26.2 (ii), (iii), which have the same frequency and amplitude as the waveform in Fig. 26.2 (i). Now curves of the shape of Fig. 26.2 (ii), (iii) can be analysed mathematically into the sum of a number of *simple harmonic* curves, whose frequencies are multiples of f_0 , the frequency of the original waveform; the amplitudes of these curves diminish as the frequency increases. Fig. 26.2 (iv), for example, might be an analysis of a curve similar to Fig. 26.2 (iii), corresponding to a note on a piano. The ear is able to detect simple harmonic waves (p. 595), and thus registers the presence of notes of frequencies $2f_0$ and $3f_0$, in addition to f_0 , when the note is sounded on the piano. The amplitude of the curve corresponding to f_0 is greatest,

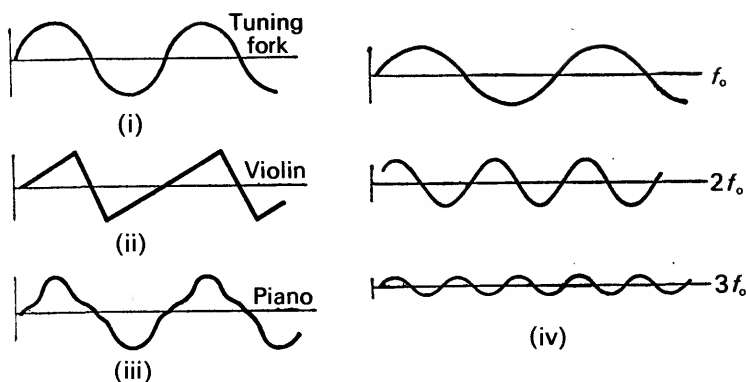


FIG. 26.2. Wave-forms of notes.

Fig. 26.2 (iv), and the note of frequency f_0 is heard predominantly because the intensity is proportional to the square of the amplitude (p. 607). In the background, however, are the notes of frequencies $2f_0$, $3f_0$, which are called the *overtones*. The frequency f_0 is called the *fundamental*.

As the waveform of the same note is different when it is obtained from different instruments, it follows that the analysis of each will differ; for example, the waveform of a note of frequency f_0 from a violin may contain overtones of frequencies $2f_0$, $4f_0$, $6f_0$. The musical “background” to the fundamental note is therefore different when it is sounded on different instruments, and hence *the overtones present in a note determine its quality or timbre*.

A *harmonic* is the name given to a note whose frequency is a simple multiple of the fundamental frequency f_0 . The latter is thus termed the “first harmonic”; a note of frequency $2f_0$ is called the “second harmonic”, and so on. Certain harmonics of a note may be absent from its overtones; for example, the only possible notes obtained from an organ-pipe closed at one end are f_0 , $3f_0$, $5f_0$, $7f_0$, and so on (p. 647).

Helmholtz Resonators

HELMHOLTZ, one of the greatest scientists of the nineteenth century, devised a simple method of detecting the overtones accompanying the fundamental note. He used vessels, P, Q, of different sizes, containing air which “responded” or *resonated* (see p. 653) to a note of a particular frequency, Fig. 26.3. When a sound wave entered a small cavity or neck

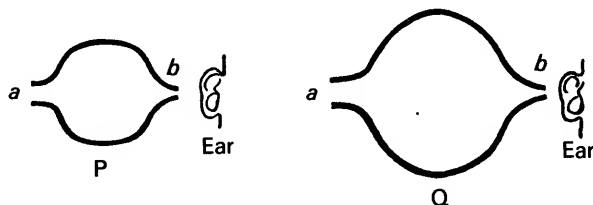


FIG. 26.3. Helmholtz resonators.

Thus the motion of the air in the neck is simple harmonic, and the period T is given by

$$T = 2\pi \sqrt{\frac{mV}{\gamma p a^2}}.$$

Hence the frequency, f , is given by

$$f = \frac{1}{T} = \frac{1}{2\pi} \sqrt{\frac{\gamma p a^2}{mV}} \quad \dots \dots \dots (iii)$$

The velocity of sound, v , is given by $v = \sqrt{\gamma p / \rho}$, where ρ is the density of the air (p. 624), or $\gamma p = v^2 \rho$. If l is the length of the neck, the mass $m = al\rho$. Thus the frequency f can also be expressed by

$$f = \frac{1}{2\pi} \sqrt{\frac{v^2 \rho a^2}{al\rho V}} = \frac{v}{2\pi} \sqrt{\frac{a}{lV}} \quad \dots \dots \dots (iv)$$

From these formulae for f , it follows that

$$f^2 V = \text{constant}.$$

The adiabatic changes at the neck are not perfect, and this result is thus only approximately true. In practice, the law more nearly obeyed is that given by

$$f^2 (V + c) = \text{constant}.$$

where c is a "correction" to V .

Experiment. In an experiment to verify the law, tuning-forks of known frequency, a bottle with a narrow neck, and a pipette and burette, are required. Water is run slowly into the bottle until resonance is obtained with the lowest note, for example. The volume of air V which is resonating is then found by subtracting the volume of the bottle below the neck, determined in a preliminary experiment, from the water run in. This is repeated for the various forks, and a graph of V is plotted against $1/f^2$. A straight line passing close to the origin is obtained, thus showing that $V + c = d/f^2$, where d is a constant, or $f^2 (V + c) = \text{constant}$.

PROPERTIES OF SOUND WAVES

Reflection

Like light waves, sound waves are reflected from a plane surface so that the angle of incidence is equal to the angle of reflection. This can be demonstrated by placing a tube T_1 in front of a plane surface AB and blowing a whistle gently at S , Fig. 26.5. Another tube T_2 , directed towards N , is placed on the other side of the normal NQ , and moved until a sensitive flame (see p. 650), or a microphone connected to a cathode-ray tube, is considerably affected at R , showing that the reflected wave passes along NR . It will then be found that angle $RNQ = \text{angle } SNQ$.

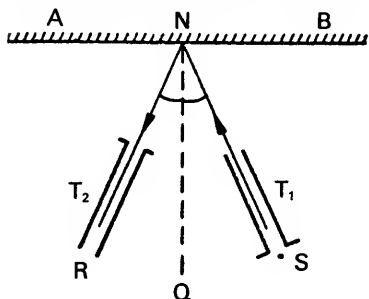


FIG. 26.5. Reflection of sound.

It can also be demonstrated that sound waves come to a focus when they are incident on a curved concave mirror. A surface shaped like a parabola reflects sound waves to

long distances if the source of sound is placed at its focus (see also p. 404). The famous whispering gallery of St. Paul's is a circular-shaped chamber whose walls repeatedly reflect sound waves round the gallery, so that a person talking quietly at one end can be heard distinctly at the other end.

Acoustics of Rooms. Reverberation

A concert-hall, lecture-room, or a broadcasting studio requires special design to be acoustically effective. The technical problems concerned were first investigated in 1906 by SABINE in America, who was consulted about a hall in which it was difficult for an audience to hear the lecturer.

Generally, an audience in a hall hears sound from different directions at different times. They hear (a) sound *directly* from the speaker or orchestra, as the case may be, (b) sound from *echoes* produced by walls and ceilings, (c) sound *diffused* from the walls and ceilings and other objects present. The echoes are due to regular reflection at a plane surface (p. 391), but the diffused sound is scattered in different directions and reflection takes place repeatedly at other surfaces. When reflection occurs some energy is absorbed from the sound wave, and after a time the sound diminishes below the level at which it can be heard. The perseverance of the sound after the source ceases is known as **reverberation**. In the case of the hall investigated by Sabine the time of reverberation was about $5\frac{1}{2}$ seconds, and the sound due to the first syllable of a speaker thus overlapped the sound due to the next dozen or so syllables, making the speech difficult to comprehend. The quality of a sound depends on the time of reverberation. If the time is very short, for example 0.5 second, the music from an orchestra sounds thin or lifeless; if the time too long the music sounds muffled. The reverberation time at a B.B.C. concert-hall used for orchestral performances is about $1\frac{3}{4}$ seconds, whereas the reverberation time for a dance-band studio is about 1 second.

Sabine's Investigations. Absorptive Power

Sabine found that the time T of reverberation depended on the volume V of the room, its surface area A , and the *absorptive power*, a , of the surfaces. The time T is given approximately by

$$T = \frac{kV}{aA},$$

where k is a constant. In general some sound is absorbed and the rest is reflected; if too much sound is reflected T is large. If many thick curtains are present in the room too much sound is absorbed and T is small.

Sabine chose the absorptive power of unit area of an open window as the unit, since this is a perfect absorber. On this basis the absorptive power of a person in an audience, or of thick carpets and rugs, is 0.5, linoleum has an absorptive power of 0.12, and polished wood and glass have an absorptive power of 0.01. The absorptive power of a material depends on its pores to a large extent; this is shown by the fact that an

unpainted brick has a high absorptive power, whereas the painted brick has a low absorptive power.

From Sabine's formula for T it follows that the time of reverberation can be shortened by having more spectators in the hall concerned, or by using felt materials to line some of the walls or ceiling. The seats in an acoustically-designed lecture-room have plush cushions at their backs to act as an absorbent of sound when the room is not full. B.B.C. studios used for plays or news talks should have zero reverberation time, as clarity is all-important, and the studios are built from special plaster or cork panels which absorb the sound completely. The structure of a room also affects the acoustics. Rooms with large curved surfaces tend to focus echoes at certain places, which is unpleasant aurally to the audience, and a huge curtain was formerly hung from the roof of the Albert Hall to obscure the dome at orchestral concerts.

Refraction

Sound waves can be refracted as well as reflected. TYNDALL placed a watch in front of a balloon filled with carbon dioxide, which is heavier than air, and found that the sound was heard at a definite place on the other side of the balloon. The sound waves thus converged to a focus on the other side of the balloon, which therefore has the same effect on sound waves as a convex lens has on light waves (see Fig. 28.12, p. 684). If the balloon is filled with hydrogen, which is lighter than air, the sound waves diverge on passing through the balloon. The latter thus acts similarly to a concave lens when light waves are incident on it (see p. 683).

The refraction of sound explains why sounds are easier to hear at night than during day-time. In the latter case the upper layers of air are colder than the layers near the earth. Now sound travels faster the higher the temperature (see 624), and sound waves are hence refracted in a direction away from the earth. The intensity of the sound waves thus diminish. At night-time, however, the layers of air near the earth are colder than those higher up, and hence sound waves are now refracted towards the earth, with a consequent increase in intensity.

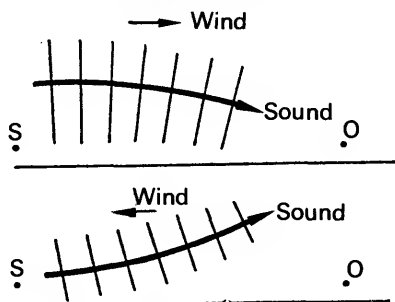


FIG. 26.6. Refraction of sound.

For a similar reason, a distant observer O hears a sound from a source S more easily when the wind is blowing towards him than away from him, Fig. 26.6. When the wind is blowing towards O , the bottom of the

sound wavefront is moving slower than the upper part, and hence the wavefronts veer towards the observer, who therefore hears the sound easily. When the wind is blowing in the opposite direction the reverse is the case, and the wavefronts veer upwards away from the ground and O. The sound intensity thus diminishes. This phenomenon is hence another example of the refraction of sound.

Interference of Sound Waves

Besides reflection and refraction, sound waves can also exhibit the phenomenon of *interference*, whose principles we shall now discuss.

Suppose two sources of sound, A, B, have exactly the same frequency and amplitude of vibration, and that their vibrations are always in phase with each other, Fig. 26.7. Such sources are called “coherent” sources. *Their combined effect at a point is obtained by adding algebraically the displacements at the point due to the sources individually*; this is known as the *Principle of Superposition*. Thus their resultant effect at X, for example, is the algebraic sum of the vibrations at X due to the source A alone and the vibrations at X due to the source B alone. If X is equidistant from A and B, the vibrations at X due to the two sources are *always* in phase as (i) the distance AX travelled by the wave originating at A is equal to the distance BX travelled by the wave originating at B, (ii) the sources A, B are assumed to have the same frequency and to be always in phase with each other. Fig. 26.8 (i), (ii) illustrate the vibrations at X due to A, B, which have the same amplitude. The resultant vibration at X is obtained by adding the two curves, and has an amplitude double that of either curve and a frequency the same as either, Fig. 26.8 (iii). Now the energy of a vibrating source is proportional to the square of its amplitude (p. 607). Consequently the sound energy at X is four times that due to A or B alone, and a loud sound is thus heard at X. As A and B are coherent sources, the loud sound is *permanent*.

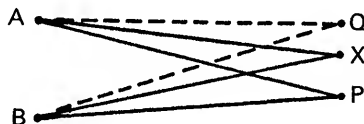


FIG. 26.7. Interference of sound.

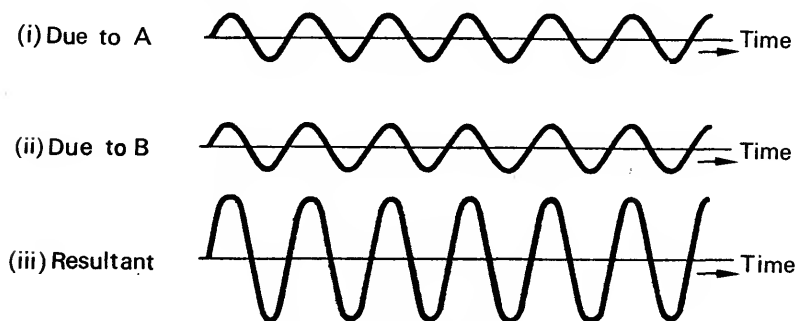


FIG. 26.8. Vibrations at X.

If Q is a point such that BQ is greater than AQ by a whole number of wavelengths (Fig. 26.7), the vibration at Q due to A is in phase with the vibration there due to B (see p. 593). A permanent loud sound is then obtained at Q . Thus a permanent loud sound is obtained at any point Y if the path difference, $BY - AY$, is given by

$$BY - AY = m\lambda,$$

where λ is the wavelength of the sources A , B , and m is an integer.

Destructive Interference

Consider now a point P in Fig. 26.7 whose distance from B is half a wavelength longer than its distance from A , i.e., $AP - BP = \lambda/2$. The vibration at P due to B will then be 180° out of phase with the vibration there to A (see p. 586), Fig. 26.9 (i), (ii). The resultant effect at P is thus zero, as the displacements at any instant are equal and opposite to each other, Fig. 26.9 (iii). No sound is therefore heard at P , and the permanent silence is said to be due to "destructive interference" between the sound waves from A and B .

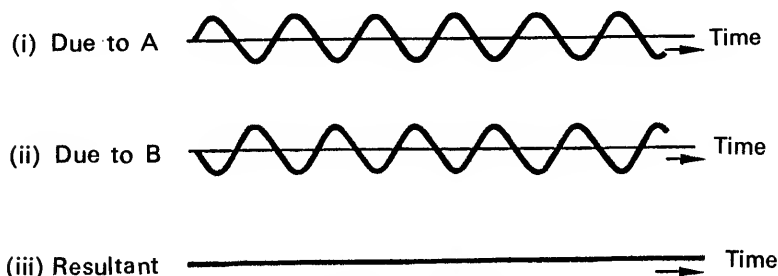


FIG. 26.9. Vibrations at P .

If the path difference, $AP - BP$, were $3\lambda/2$ or $5\lambda/2$, instead of $\lambda/2$, permanent silence would also exist at P as the vibrations there due to A , B would again be 180° out of phase. Summarising, then,

silence occurs if the path-difference is an odd number of half wavelengths, and

a loud sound occurs if the path-difference is a whole number of wavelengths

The total sound energy in all the positions of loud sound discussed above is equal to the total sound energy of the two sources A , B , from the principle of the conservation of energy. The extra sound at the positions of loud sound thus makes up for the absent sound in the positions of silence.

Quincke's Tube. Measurement of Velocity of Sound in a Tube

QUINCKE devised a simple method of obtaining permanent interference between two sound waves. He used a closed tube $SAEB$ which had openings at S , E , and placed a source of sound at S , Fig. 26.10. A wave then travelled in the direction SAE round the tube, while another wave travelled in the opposite direction SBE ; and since these waves

are due to the same source, S, they always set out in phase, i.e., they are coherent.

Like a trombone, one side, B, of the tube can be pulled out, thus making SAE, SBE of different lengths. When SAE and SBE are equal in length an observer at E hears a loud sound, since the paths of the two waves are then equal. As B is pulled out the sound dies away and becomes a minimum when the path difference, $SBE - SAE$, is $\lambda/2$, where λ is the wavelength. In this case the two waves arrive 180° out of phase (p. 617). If the tube is pulled out farther, the sound increases in loudness to a maximum; the path difference is then λ . If k is the distance moved from one position of minimum sound, MN say, to the next position of minimum sound, PQ say, then $2k = \lambda$, Fig. 26.10. Thus the wavelength of the sound can be simply obtained by measuring k .

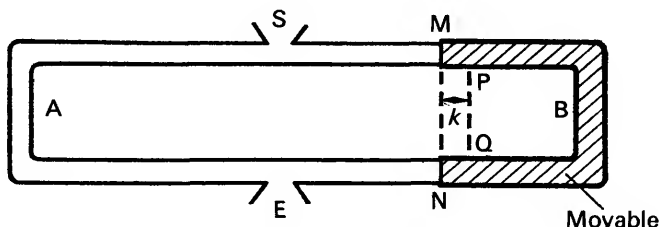


FIG. 26.10. Quincke's tube.

The velocity of sound in the tube is given by $V = f\lambda$, where f is the frequency of the source S, and thus V can be found when a source of known frequency is used. In a particular experiment with Quincke's tube, the tube B was moved a distance 4.28 cm between successive minima of sound, and the frequency of the source was 4000 Hz.

Thus $\lambda = 2 \times 4.28$ cm,

and $V = f\lambda = 4000 \times 2 \times 4.28 = 34240 \text{ cm s}^{-1} = 342.4 \text{ m s}^{-1}$

It can be seen that, unlike reflection and refraction, the phenomenon of interference can be utilised to measure the wavelength of sound waves. We shall see later that interference is also utilised to measure the wavelength of light waves (p. 689).

Velocity of Sound in Free Air. Hebb's Method

In 1905 HEBB performed an accurate experiment to measure the velocity of sound in free air which utilised a method of interference. He carried out his experiment in a large hall to eliminate the effect of wind, and obtained the temperature of the air by placing thermometers at different parts of the room. Two parabolic reflectors, R_1 , R_2 , are placed at each end of the hall, and microphones, M_1 , M_2 , are positioned at the respective foci, S_1 , S_2 to receive sound reflected from R_1 , R_2 , Fig. 26.11. By means of a transformer, the currents in the microphones are induced into a telephone earpiece P, so that the resultant effect of the sound waves received by M_1 , M_2 respectively can be heard.

A source of sound of known constant frequency is placed at the focus S_1 . The sound waves are reflected from R_1 in a parallel direction (p. 613), and travel to R_2 where they are reflected to the focus S_2 and

received by M_2 . The microphone M_1 receives sound waves directly from the source, and hence the sound heard in the telephone earpiece is due to the resultant effect of two coherent sources. With the source

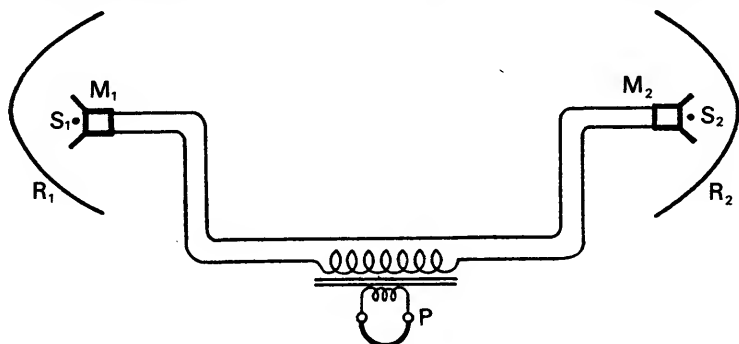


FIG. 26.11. Hebb's method.

and microphone maintained at its focus S_1 , R_1 is moved along its axis in one direction. The positions of R_1 are noted when minima of sound are heard; and since the distance between successive minima corresponds to one wavelength, λ , the velocity of sound can be calculated from the relation $V = f\lambda$, as f and λ are known.

Other Velocity of Sound Determinations

The velocity of sound in *air* has been determined by many scientists. One of the first accurate determinations was carried out in 1738 by French scientists, who observed the time between the flash and the hearing of a cannon report about 30 km away. Their results confirmed that the velocity of sound increased as the temperature of the air increased (p. 624), and they obtained the result of 362 metres per second for the velocity at 0°C . Similar experiments were carried out by French scientists in 1822. In 1844 experiments carried out in the Tyrol district, several thousand metres above sea-level, showed that the velocity of sound was independent of the pressure of the air (p. 624).

REGNAULT, the eminent French experimental scientist of the nineteenth century, carried out an accurate series of measurements on the velocity of sound in 1864. Guns were fired at one place, breaking an electrical circuit automatically, and the arrival of the sound at a distant place was recorded by a second electrical circuit. Both circuits actuated a pen or style pressing against a drum rotating at a steady speed round its axis, which is known as a *chronograph*. Thus marks were made on the drum at the instant the sound occurred and the instant it was received. The small interval corresponding to the distance between the marks was determined from a wavy trace made on the drum by a style attached to an electrically-maintained tuning-fork whose frequency was known, and the speed of sound was thus calculated.

The velocity of sound in *water* was first accurately determined in 1826. The experiment was carried out by immersing a bell in the Lake of Geneva, and arranging to fire gunpowder at the instant the bell was

struck. Miles away, the interval was recorded between the flash and the later arrival of the sound in the water, and the velocity was then calculated. This and other experiments have shown that the velocity in water is about 1435 m s^{-1} , more than four times the speed in air.

An objection to all these methods of determining velocity is the unknown time lag between the receipt of the sound by an observer and his recording of the sound. The observer has, as it were, a "personal equation" which must be taken into account to determine the true time of travel of the sound. In Hebb's method, however, which utilises interference, no such personal equation enters into the considerations, which is an advantage of the method.

Beats

If two notes of nearly equal frequency are sounded together, a periodic rise and fall in intensity can be heard. This is known as the phenomenon of *beats*, and the frequency of the beats is the number of intense sounds heard per second.

Consider a layer of air some distance away from two pure notes of nearly equal frequency, say 48 and 56 Hz respectively, which are sounding. The variation of the displacement, y_1 , of the layer due to one fork alone is shown in Fig. 26.12 (i); the variation of the displacement y_2 ,

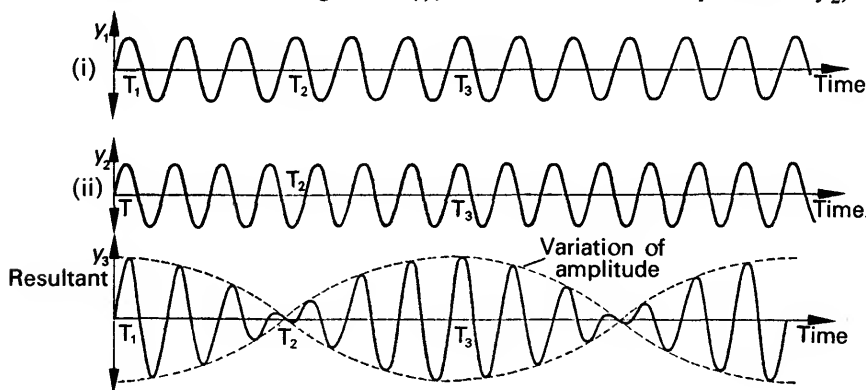


FIG. 26.12. Beats (*not to scale*).

of the layer due to the second fork alone is shown in Fig. 26.12 (ii). According to the Principle of Superposition (p. 588), the variation of the resultant displacement, y , of the layer is the algebraic sum of the two curves, which varies in amplitude in the way shown in Fig. 26.12 (iii). To understand the variation of y , suppose that the displacements y_1 , y_2 are in phase at some instant T_1 , Fig. 26.12. Since the frequency of the curve in Fig. 26.12 (i) is 48 cycles per sec the variation y_1 undergoes 3 complete cycles in $\frac{1}{16}$ th second; in the same time, the variation y_2 undergoes $3\frac{1}{2}$ cycles, since its frequency is 56 cycles per second. Thus y_1 and y_2 are 180° out of phase with each other at this instant, and their resultant y is then a minimum at some instant T_2 . Thus T_1T_2 represents $\frac{1}{16}$ th of a second in Fig. 26.12 (iii). In $\frac{1}{8}$ th of a second from T_1 , y_1 has undergone 6 complete cycles and y_2 has undergone 7 complete cycles. The two waves are

hence in phase again at T_3 , where $T_1 T_3$ represents $\frac{1}{8}$ th of a second, and their resultant at their instant is again a maximum, Fig. 26.12 (iii). In this way it can be seen that a loud sound is heard after every $\frac{1}{8}$ second, and thus the beat frequency is 8 cycles per second. This is the difference between the frequencies, 48, 56, of the two notes, and it is shown soon that *the beat frequency is always equal to the difference of the two nearly equal frequencies*.

It can now be seen that beats are a phenomenon of repeated interference. Unlike the cases in sound previously considered, however, the two sources are not coherent ones.

Beat Frequency Formula

Suppose two sounding tuning-forks have frequencies f_1, f_2 cycles per second close to each other. At some instant of time the displacement of a particular layer of air near the ear due to each fork will be a maximum to the right. The resultant displacement is then a maximum, and a loud sound or beat is heard. After this, the vibrations of air due to each fork go out of phase, and t seconds later the displacement due to each fork is again a maximum to the right, so that a loud sound or beat is heard again. One fork has then made exactly one cycle more than the other. But the number of cycles made by each fork in t seconds is $f_1 t$ and $f_2 t$ respectively. Assuming f_1 is greater than f_2 ,

$$\therefore f_1 t - f_2 t = 1$$

$$\therefore f_1 - f_2 = \frac{1}{t}$$

Now 1 beat has been made in t seconds, so that $1/t$ is the number of beats per second or beat frequency.

$$\therefore f_1 - f_2 = \text{beat frequency.}$$

Mathematical derivation of beat frequency. Suppose y_1, y_2 are the displacements of a given layer of air due to two tuning-forks of frequencies f_1, f_2 respectively. If the amplitudes of each vibration are equal to a , then $y_1 = a \sin \omega_1 t, y_2 = a \sin (\omega_2 t + \theta)$, where $\omega_1 = 2\pi f_1, \omega_2 = 2\pi f_2$, and θ is the constant phase angle between the two variations.

$$\therefore y = y_1 + y_2 = a [\sin \omega_1 t + \sin (\omega_2 t + \theta)]$$

$$\therefore y = 2a \sin \left(\frac{\omega_1 + \omega_2}{2} t + \frac{\theta}{2} \right) \cdot \cos \left(\frac{\omega_1 - \omega_2}{2} t - \frac{\theta}{2} \right)$$

$$\therefore y = A \sin \left(\frac{\omega_1 + \omega_2}{2} t + \frac{\theta}{2} \right)$$

$$\text{where } A = 2a \cos \left(\frac{\omega_1 - \omega_2}{2} t - \frac{\theta}{2} \right).$$

We can regard A as the *amplitude* of the variation of y . The intensity of the resultant note is proportional to A^2 , the square of the amplitude (p. 607) and

$$A^2 = 4a^2 \cos^2 \left(\frac{\omega_1 - \omega_2}{2} t - \frac{\theta}{2} \right) = 2a^2 [1 + \cos (\omega_1 - \omega_2) t - \theta]$$

since $2 \cos^2 \alpha = 1 + \cos 2\alpha$. It then follows that the intensity varies at a frequency f given by

$$2\pi f = \omega_1 - \omega_2.$$

But $\omega_1 = 2\pi f_1$, $\omega_2 = 2\pi f_2$.

$$\therefore 2\pi f = 2\pi f_1 - 2\pi f_2$$

$$\therefore f = f_1 - f_2$$

The frequency f of the beats is thus equal to the difference of the frequencies.

Uses of Beats

The phenomenon of beats can be used to measure the unknown frequency, f_1 , of a note. For this purpose a note of known frequency f_2 is used to provide beats with the unknown note, and the frequency f of the beats is obtained by counting the number made in a given time. Since f is the difference between f_2 and f_1 , it follows that $f_1 = f_2 - f$, or $f_1 = f_2 + f$. Thus suppose $f_2 = 1000$ Hz, and the number of beats per second made with a tuning-fork of unknown frequency f_1 is 4. Then $f_1 = 1004$ or 996 Hz.

To decide which value of f_1 is correct, the end of the tuning-fork prong is loaded with a small piece of plasticine, which diminishes the frequency a little, and the two notes are sounded again. If the beats are *increased*, a little thought indicates that the frequency of the note must have been originally 996 Hz. If the beats are decreased, the frequency of the note must have been originally 1004 Hz. The tuning-fork must not be overloaded, as the frequency may decrease, if it was 1004 Hz, to a frequency such as 995 Hz, in which case the significance of the beats can be wrongly interpreted.

Beats are also used to "tune" an instrument to a given note. As the instrument note approaches the given note, beats are heard, and the instrument can be regarded as "tuned" when the beats are occurring at a very slow rate.

Velocity of Sound in a Medium

When a sound wave travels in a medium, such as a gas, a liquid, or a solid, the particles in the medium are subjected to varying stresses, with resulting strains (p. 585). The velocity of a sound wave is thus partly governed by the *modulus of elasticity*, E , of the medium, which is defined by the relation

$$E = \frac{\text{stress}}{\text{strain}} = \frac{\text{force per unit area}}{\text{change in length (or volume) / original length (or volume)}} \quad (i)$$

The velocity, V , also depends on the density, ρ , of the medium, and it can be shown that

$$V = \sqrt{\frac{E}{\rho}} \quad (1)$$

When E is in newton per metre² (N m⁻²) and ρ in kg m⁻³, then V is in metre per second (m s⁻¹). The relation (1) was first derived by Newton.

For a solid, E is Young's modulus of elasticity. The magnitude of E

$$V = \sqrt{\frac{E}{\rho}} = \sqrt{\frac{2 \times 10^{11}}{7800}} = 5060 \text{ m s}^{-1}$$
$$V = \sqrt{\frac{2.04 \times 10^9}{1000}} = 1430 \text{ m s}^{-1}$$
$$\frac{\text{ML}}{\text{T}^2\text{I}^2} \text{ or } \text{ML}^{-1}\text{T}^{-2}$$
$$\therefore \text{LT}^{-1} = (\text{ML}^{-1}\text{T}^{-2})^x \times (\text{ML}^{-3})^y$$
$$x + y = 0 \quad \text{. (ii)}$$

$$-x - 3y = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (iii)$$

$$-2x = -1 \quad \therefore \therefore \therefore \therefore \therefore \quad \text{(iv)}$$

$$V = kE^{\frac{1}{2}} \rho^{-\frac{1}{2}}$$

$$\therefore V = k \sqrt{\frac{E}{\rho}}$$

a rigid proof of the formula by calculus shows that $k = 1$ since $V = \sqrt{\frac{E}{\rho}}$.

The velocity of sound in a gas is also given by $V = \sqrt{\frac{E}{\rho}}$ where E is

$V = \sqrt{\frac{p}{\rho}}$ and as the density, ρ , of air is 1.29 kg per m³ at S.T.P. and

Ultrasonics

There are sound waves of higher frequency than 20000 Hz, which are inaudible to a human being. These are known as *ultrasonics*; and since $\text{velocity} = \text{wavelength} \times \text{frequency}$, ultrasonics have short wavelengths compared with sound waves in the audio-frequency range.

In recent years ultrasonics have been utilised for a variety of industrial purposes. They are used on board coasting vessels for depth sounding, the time taken by the wave to reach the bottom of the sea from the surface and back being determined. Ultrasonics are also used to kill bacteria in liquids, and they are used extensively to locate faults and cracks in metal castings, following a method similar to that of radar. Ultrasonic waves are sent into the metal under investigation, and the beam reflected from the fault is picked up on a cathode-ray tube screen together with the reflection from the other end of the metal. The position of the fault can then easily be located.

Production of Ultrasonics

In 1881 CURIE discovered that a thin plate of quartz increased or decreased in length if an electrical battery was connected to its opposite faces. By correctly cutting the plate, the expansion or contraction could be made to occur along the axis of the faces to which the battery was applied. When an alternating voltage of ultrasonic frequency was connected to the faces of such a crystal the faces vibrated at the same frequency, and thus ultrasonic sound waves were produced.

Another method of producing ultrasonics is to place an iron or nickel rod inside a solenoid carrying an alternating current of ultrasonic frequency. Since the length of a magnetic specimen increases slightly when it is magnetised, ultrasonic sound waves are produced by the vibrations of the rod.

EXAMPLES

1. How does the velocity of sound in a medium depend upon the elasticity and density? Illustrate your answer by reference to the case of air and of a long metal rod. The velocity of sound in air being 330.0 m s^{-1} at 0° C and the coefficient of expansion $1/273$ per degree, find the change in velocity per degree Centigrade rise of temperature. (*L.*)

First part. The velocity of sound, V , is given by $V = \sqrt{E/\rho}$, where E is the modulus of elasticity of the medium and ρ is its density. In the case of air, a gas, E represents the bulk of modulus of the air under adiabatic conditions, and $E = \gamma p$ (see p. 624). Thus $V = \sqrt{\gamma p/\rho}$ for air.

For a long metal rod, E is Young's modulus for the metal, assuming the sound travels along the length of the rod.

Second part. Since the coefficient of expansion is $1/273$ per degree Centigrade, the absolute temperature corresponding to $t^\circ \text{ C}$ is given by $(273 + t)$. The velocity of sound in a gas is proportional to the square root of its absolute temperature, and hence

$$\frac{V}{V_0} = \sqrt{\frac{274}{273}},$$

where V is the velocity at 1°C and V_0 is the velocity at 0°C .

$$\therefore V = V_0 \sqrt{\frac{274}{273}} = 330 \times \sqrt{\frac{274}{273}} = 330.6 \text{ m s}^{-1}$$

$$\therefore \text{change in velocity} = 0.6 \text{ m s}^{-1}$$

2. How would you find by experiment the velocity of sound in air? Calculate the velocity of sound in air in metre second $^{-1}$ at 100°C if the density of air at S.T.P. is 1.29 kg m^{-3} , the density of mercury at 0°C 13600 kg m^{-3} , the specific heat capacity of air at constant pressure 1.02 , and the specific heat capacity of air at constant volume 0.72 , in $\text{kJ kg}^{-1}\text{K}^{-1}$. (L)

First part. See Hebb's method, p. 618 or p. 587.

Second part. The velocity of sound in air is given by

$$V = \sqrt{\frac{\gamma p}{\rho}}$$

with the usual notation. The density, ρ , of air is 1.29 kg m^{-3} . The pressure p is given by

$$\begin{aligned} p &= h \rho g \\ &= 0.76 \times 13600 \times 9.8 \text{ N m}^{-2} \end{aligned}$$

since S.T.P. denotes 76 cm mercury pressure and 0°C . Also,

$$\gamma = \frac{C_p}{C_v} = \frac{1.02}{0.72}$$

$$\therefore V = \sqrt{\frac{1.02 \times 0.76 \times 13600 \times 9.8}{0.72 \times 1.293}},$$

where V is the velocity at 0°C .

But

$$\text{velocity} \propto \sqrt{T},$$

where T is the absolute temperature of the air. Thus if V' is the velocity at 100°C ,

$$\frac{V'}{V} = \sqrt{\frac{273 + 100}{273}} = \sqrt{\frac{373}{273}}$$

$$\therefore V' = \sqrt{\frac{373}{273}} V = \sqrt{\frac{373 \times 1.02 \times 0.76 \times 13600 \times 9.8}{273 \times 0.72 \times 1.293}}$$

$$\therefore V' = 388 \text{ m s}^{-1}$$

3. State briefly how you would show by experiment that the characteristics of the transmission of sound are such that (a) a finite time is necessary for transmission, (b) a material medium is necessary for propagation, (c) the disturbance may be reflected and refracted. The wavelength of the note emitted by a tuning-fork, frequency 512 Hz , in air at 17°C is 66.5 cm . If the density of air at S.T.P. is 1.293 kg m^{-3} , calculate the ratio of the two specific heat capacities of air. Assume that the density of mercury is 13600 kg m^{-3} . (N)

First part. See text.

Second part. Since $V = f\lambda$, the velocity of sound at 17°C is given by

$$V = 512 \times 0.665 \text{ m s}^{-1} \quad . \quad . \quad . \quad (i)$$

Now
$$\frac{V_0}{V} = \sqrt{\frac{273}{290}}$$

where V_0 is the velocity at 0°C , since the velocity is proportional to the square root of the absolute temperature.

$$\therefore V_0 = \sqrt{\frac{273}{290}} \times V = \sqrt{\frac{273}{290}} \times 512 \times 0.665 \quad \quad (ii)$$

But
$$V_0 = \sqrt{\frac{\gamma p}{\rho}},$$

where $p = 0.76 \text{ m of mercury} = 0.76 \times 13600 \times 9.8 \text{ N m}^{-2}$, and $\rho = 1.293 \text{ kg m}^{-3}$.

$$\begin{aligned} \therefore \gamma &= \frac{V_0^2 \times \rho}{p} \\ &= \frac{272 \times 512^2 \times 0.665 \times 1.293}{290 \times 0.76 \times 13600 \times 9.8} \\ &= 1.39 \end{aligned}$$

Doppler Effect

The whistle of a train or a jet aeroplane appears to increase in pitch as it approaches a stationary observer; as the moving object passes the observer, the pitch changes and becomes lowered. The apparent alteration in frequency was first predicted by DOPPLER in 1845, who stated that a change of frequency of the wave-motion should be observed when a source of sound or light was moving, and it is accordingly known as the *Doppler effect*.

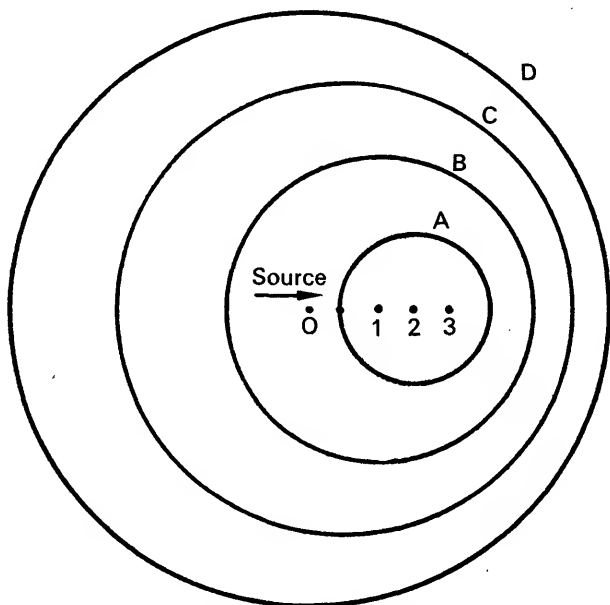


FIG. 26.13. Doppler effect.

The Doppler effect occurs whenever there is a *relative velocity* between the source of sound or light and an observer. In light, this effect was observed when measurements were taken of the wavelength of the colour of a moving star; they showed a marked variation. In sound, the Doppler effect can be demonstrated by placing a whistle in the end of a long piece of rubber tubing, and whirling the tube in a horizontal circle above the head while blowing the whistle. The open end of the tube acts as a moving source of sound, and an observer hears a rise and fall in pitch as the end approaches and recedes from him.

A complete calculation of the apparent frequency in particular cases is given shortly, but Fig. 26.13 illustrates how the change of wavelengths, and hence frequency, occurs when a source of sound is moving towards a stationary observer. At a certain instant the position of the moving source is at 4. At four successive seconds *before* this instant the source had been at the positions 3, 2, 1, 0 respectively. If V is the velocity of sound, the wavefront from the source when in the position 3 reaches the surface A of a sphere of radius V and centre 3 when the source is just at 4. In the same way, the wavefront from the source when it was in the position 2 reaches the surface B of a sphere of radius $2V$ and centre 2. The wavefront C corresponds to the source when it was in the position 1, and the wavefront D to the source when it was in the position 0. Thus if the observer is on the right of the source S, he receives wavefronts which are relatively more crowded together than if S were stationary; the frequency of S thus appears to increase. When the observer is on the left of S, in which case the source is moving away from him, the wavefronts are farther apart than if S were stationary, and hence the observer receives correspondingly fewer waves per second. The apparent frequency is thus lowered.

Calculation of Apparent Frequency

Suppose V is the velocity of sound in air, u_s is the velocity of the source of sound S, u_o is the velocity of an observer O, and f is the true frequency of the source.

(i) *Source moving towards stationary observer.* If the source S were stationary, the f waves sent out in one second towards the observer O

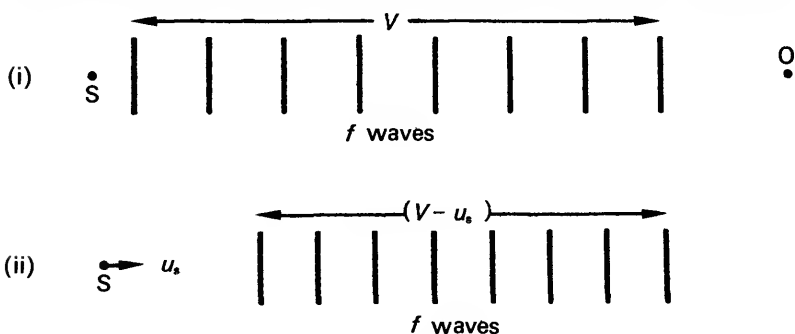


FIG. 26.14. Source moving towards stationary observer.

would occupy a distance V , and the wave length would be V/f , Fig. 26.14.

(i). If S moves with a velocity u_s towards O , however, the f waves sent out occupy a distance $(V - u_s)$, because S has moved a distance u_s towards O in 1 sec, Fig. 26.14 (ii). Thus the wavelength λ' of the waves reaching O is now $(V - u_s)/f$.

But velocity of sound waves = V .

$$\therefore \text{apparent frequency, } f' = \frac{V}{\lambda'} = \frac{V}{(V - u_s)/f}$$

$$\therefore f' = \frac{V}{V - u_s} f \quad . \quad . \quad . \quad (3)$$

Since $(V - u_s)$ is less than V , f' is greater than f ; the apparent frequency thus appears to increase when a source is moving towards an observer.

(ii) *Source moving away from stationary observer.* In this case the f waves sent out towards O in 1 sec occupy a distance $(V + u_s)$, Fig. 26.15.

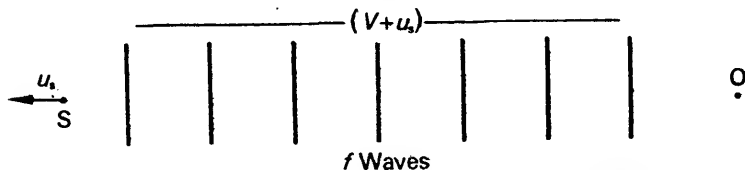


FIG. 26.15. Source moving away from stationary observer.

The wavelength λ' of the waves reaching O is thus $(V + u_s)/f$, and hence the apparent frequency f' is given by

$$f' = \frac{V}{\lambda'} = \frac{V}{(V + u_s)/f}$$

$$\therefore f' = \frac{V}{V + u_s} \cdot f \quad . \quad . \quad . \quad (4)$$

Since $(V + u_s)$ is greater than V , f' is less than f , and hence the apparent frequency decreases when a source moves away from an observer.

(iii) *Source stationary, and observer moving towards it.* Since the source is stationary, the f waves sent out by S towards the moving observer O occupies a distance V , Fig. 26.16. The wavelength of the waves reaching O is hence V/f , and thus unlike the cases already considered, the wavelength is unaltered.

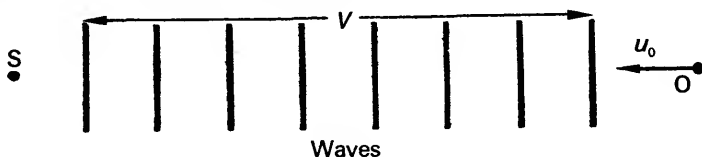


FIG. 26.16. Observer moving towards stationary source.

The velocity of the sound waves relative to O is not V , however, as O is moving relative to the source. The velocity of the sound waves relative

effect. Suppose the velocity of the wind is u_w , in the direction of the line SO joining the source S to the observer O. Since the air has then a velocity u_w relative to the ground, and the velocity of the sound waves relative to the air is V , the velocity of the waves relative to ground is $(V + u_w)$ if the wind is blowing in the same direction as SO. All our previous expressions for f' can now be adjusted by replacing the velocity V in it by $(V + u_w)$. If the wind is blowing in the opposite direction to SO, the velocity V must be replaced by $(V - u_w)$.

When the source is moving at an angle to the line joining the source and observer, the apparent frequency changes continuously. Suppose the source is moving along AB with a velocity v , while the observer is stationary at O, Fig. 229. At S, the component velocity of v along OS is $v \cos \theta$, and is towards O. The observer thus hears a note of higher pitch whose frequency f' is given by

$$f' = \frac{V}{V - v \cos \theta} f,$$

where V is the velocity of sound and f is the frequency of the source of sound. See equation (3), in which u_s now becomes $v \cos \theta$. When the source reaches P, Fig. 26.18, the component of v is $v \cos \alpha$ away from O, and the apparent frequency f'' is given by

$$f'' = \frac{V}{V + v \cos \alpha} f,$$

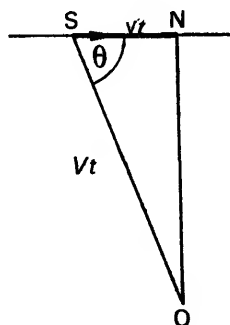


FIG. 26.19. Frequency heard when source at N.

In this case $SN = vt$ and $SO = Vt$, where t is the time-interval concerned. Thus:

$$\cos \theta = \frac{vt}{Vt} = \frac{v}{V}.$$

The frequency f' of the note heard by O when S just reaches N is hence given by

$$f' = \frac{V}{V - v \cos \theta} \cdot f = \frac{V}{V - v^2/V} \cdot f = \frac{V^2}{V^2 - v^2} \cdot f$$

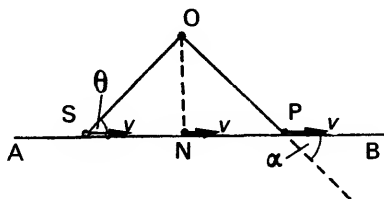


FIG. 26.18. Source direction perpendicular to observer.

from equation (4). The apparent frequency is thus lower than the frequency f of the source. When the source reaches N, the foot of the perpendicular from O to AB, the velocity v is perpendicular to ON and has thus no component towards the observer O. If the waves reach O shortly after, the observer hears a note of the same frequency f as the source.

Before the source S reaches N, however, it emits waves, travelling with a velocity V in air which reach O, Fig. 26.19. If S reaches N at the same instant as the waves reach O, the observer hears the note corresponding to the instant

when the source was at S. In this case $SN = vt$ and $SO = Vt$, where t is the time-interval concerned. Thus:

Doppler's Principle in Light

The speed of distant stars and planets has been estimated from measurements of the wavelengths of the spectrum lines which they emit. Suppose a star or planet is moving with a velocity v away from the earth and emits light of wavelength λ . If the frequency of the vibrations is f cycles per second, then f waves are emitted in one second, where $c = f\lambda$ and c is the velocity of light in *vacuo*. Owing to the velocity v , the f waves occupy a distance $(c + v)$. Thus the *apparent wavelength* λ' to an observer on the earth in line with the star's motion is

$$\lambda' = \frac{c + v}{f} = \frac{c + v}{c} \cdot \lambda = \left(1 + \frac{v}{c}\right) \lambda$$

$$\therefore \lambda' - \lambda = \text{"shift" in wavelength} = \frac{v}{c} \lambda, \quad (i)$$

$$\text{and hence } \frac{\lambda' - \lambda}{\lambda} = \text{fractional change in wavelength} = \frac{v}{c}. \quad (ii)$$

From (i), it follows that λ' is greater than λ when the star or planet is moving away from the earth, that is, there is a "shift" or displacement *towards the red*. The position of a particular wavelength in the spectrum of the star is compared with that obtained in the laboratory, and the difference in the wavelengths, $\lambda' - \lambda$, is measured. From (i), knowing λ and c , the velocity v can be calculated.

If the star is moving *towards* the earth with a velocity u , the apparent wavelength λ'' is given by

$$\lambda'' = \frac{c - u}{f} = \frac{c - u}{c} \cdot \lambda = \left(1 - \frac{u}{c}\right) \lambda.$$

$$\therefore \lambda - \lambda'' = \frac{u}{c} \lambda.$$

Since λ'' is less than λ , there is a displacement towards the blue in this case.

In measuring the speed of a star, a photograph of its spectrum is taken. The spectral lines are then compared with the same lines obtained by photographing in the laboratory an arc or spark spectrum of an element present in the star. If the former are displaced towards the red, the star is receding from the earth; if it is displaced towards the violet, the star is approaching the earth. By this method the velocities of the stars have been found to be between about 10 km s^{-1} and 300 km s^{-1} . The Doppler effect has also been used to measure the speed of rotation of the sun. Photographs are taken of the east and west edges of the sun; each contains absorption lines due to elements such as iron vaporised in the sun, and also some absorption lines due to oxygen in the earth's atmosphere. When the two photographs are put together so that the oxygen lines coincide, the iron lines in the two photographs are displaced relative to each other. In one case the edge of the sun approaches the earth, and in the other the opposite edge recedes from the earth. Measurements show a rotational speed of about 2 km s^{-1} .

Measurement of Plasma Temperature

In very hot gases or plasma, used in thermonuclear fusion experiments, the temperature is of the order of millions of degrees Celsius. At these high temperatures molecules of the glowing gas are moving away and towards the observer with very high speeds and, owing to the Doppler effect, the wavelength λ of a particular spectral line is apparently changed. One edge of the line now corresponds to an apparently increased wavelength λ_1 due to molecules moving directly towards the observer, and the other edge to an apparent decreased wavelength λ_2 due to molecules moving directly away from the observer. The line is thus observed to be *broadened*.

From our previous discussion, if v is the velocity of the molecules,

$$\lambda_1 = \frac{c + v}{c} \cdot \lambda$$

and

$$\lambda_2 = \frac{c - v}{c} \cdot \lambda$$

$$\therefore \text{breadth of line, } \lambda_1 - \lambda_2 = \frac{2v}{c} \cdot \lambda \quad (i)$$

The breadth of the line can be measured by a diffraction grating, and as λ and c are known, the velocity v can be calculated. By the kinetic theory of gases, the velocity v of the molecules is roughly the root-mean-square velocity, or $\sqrt{3RT}$, where T is the absolute temperature and R is the gas constant per gram of the gas. Consequently T can be found.

Doppler Effect and Radio Waves

A radio wave is an electromagnetic wave, like light, and travels with the same velocity, c , in free space of $3.0 \times 10^5 \text{ km s}^{-1}$. The Doppler effect with radio waves can be utilised for finding the speed of aeroplanes and satellites.

As an illustration, suppose an aircraft C sends out two radio beams at a frequency of 10^{10} Hz ; one in a forward direction, and the other in a backward direction, each beam being inclined downward at an angle of 30° to the horizontal, Fig. 26.20. A Doppler effect is obtained when the radio waves are scattered at the ground at A, B, and when the

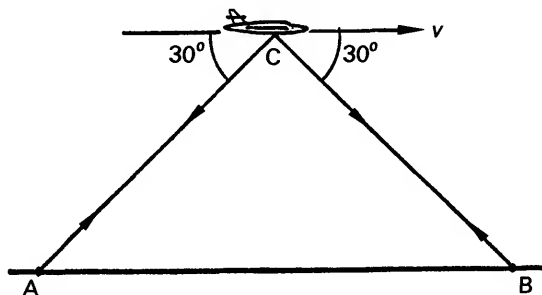


FIG. 26.20. Doppler effect and radio waves.

returning waves to C are combined, a beat frequency equal to their difference is measured. Suppose the beat frequency is 3×10^4 Hz.

If the velocity of the aircraft C is v , the velocity of radio waves is c and the frequency of the emitted beams is f , the apparent frequency f' of the waves reaching A is given by

$$f' = \frac{c}{c + v \cos \theta} \cdot f, \quad \text{. (i)}$$

where θ is 30° . The frequency f_1 of the wave received back at C from A is given by

$$f_1 = \frac{V'}{\lambda'}$$

where V' is the velocity of the wave relative to C and λ' is the wavelength of the waves reaching C. Since $V' = c - v \cos \theta$ and $\lambda' = cf'$,

$$\therefore f_1 = \frac{c - v \cos \theta}{c} f' = \frac{c - v \cos \theta}{c + v \cos \theta} \cdot f, \quad \text{. . . (ii)}$$

from (i). Similarly, the frequency f_2 of the waves received back at C from B is given by

$$f_2 = \frac{c + v \cos \theta}{c - v \cos \theta} \cdot f \quad \text{. (iii)}$$

$$\therefore \text{beat frequency at C} = f_2 - f_1 = \frac{4cv \cos \theta}{c^2 - v^2 \cos^2 \theta} \cdot f$$

Now $c = 3 \times 10^5 \text{ km s}^{-1}$, $\theta = 30^\circ$, $f_2 - f_1 = 3 \times 10^4 \text{ Hz}$, $f = 10^{10} \text{ Hz}$, and $v^2 \cos^2 \theta$ is negligible compared with c^2 .

$$\therefore 3 \times 10^4 = \frac{4cv \cos \theta}{c^2} \cdot f = \frac{4v \cos \theta}{c} \cdot f$$

$$\begin{aligned} \therefore v &= \frac{3 \times 10^4 \times 3 \times 10^5}{4 \cos 30^\circ \times 10^{10}} \\ &= 0.26 \text{ km s}^{-1} \\ &= 936 \text{ km h}^{-1} \text{ (approx.)} \end{aligned}$$

The speed of the aircraft relative to the ground is thus nearly 940 km h^{-1}

EXAMPLES

1. Obtain the formula for the Doppler effect when the source is moving with respect to a stationary observer. Give examples of the effect in sound and light. A whistle giving out 500 Hz moves away from a stationary observer in a direction towards and perpendicular to a flat wall with a velocity of 1.5 m s^{-1} . How many beats per sec will be heard by the observer? [Take the velocity of sound as 336 m s^{-1} and assume there is no wind.] (C.)

First part. See text.

Second part. The observer hears a note of apparent frequency f' from the whistle directly, and a note of apparent frequency f'' from the sound waves reflected from the wall.

Now
$$f' = \frac{V'}{\lambda'}$$

where V' is the velocity of sound in air relative to the observer and λ' is the wavelength of the waves reaching the observer. Since $V' = 336 \text{ m s}^{-1}$ and

$$\lambda' = \frac{336 + 1.5}{500} \text{ m}$$

$$\therefore f' = \frac{336 \times 500}{337.5} = 497.8 \text{ Hz}$$

The note of apparent frequency f'' is due to sound waves moving towards the observer with a velocity of 1.5 m s^{-1}

$$\begin{aligned} \therefore f'' &= \frac{V'}{\lambda'} = \frac{336}{(336 - 1.5)/500} \\ &= \frac{336 \times 500}{334.5} = 502.2 \text{ Hz} \end{aligned}$$

$$\therefore \text{beats per second} = f'' - f' = 502.2 - 497.8 = 4.4$$

2. Two observers A and B are provided with sources of sound of frequency 500. A remains stationary and B moves away from him at a velocity of 1.8 m s^{-1} . How many beats per sec are observed by A and by B, the velocity of sound being 330 m s^{-1} ? Explain the principles involved in the solution of this problem. (L.)

Beats observed by A. A hears a note of frequency 500 due to its own source of sound. He also hears a note of apparent frequency f' due to the moving source B. With the usual notation,

$$f' = \frac{V'}{\lambda'} = \frac{330}{(330 + 1.8)/500}$$

since the velocity of sound, V' , relative to A is 330 m s^{-1} and the wavelength λ' of the waves reaching him is $(330 + 1.8)/500 \text{ m}$.

$$\therefore f' = \frac{330 \times 500}{331.8} = 497.3$$

$$\therefore \text{beats observed by A} = 500 - 497.29 = 2.71 \text{ Hz.}$$

Beats observed by B. The apparent frequency f' of the sound from A is given by

$$f' = \frac{V'}{\lambda'}$$

In this case $V' =$ velocity of sound relative to B $= 330 - 1.8 = 328.2 \text{ m s}^{-1}$ and the wavelength λ' of the waves reaching B is unaltered. Since $\lambda' = 330/500 \text{ m}$, it follows that

$$f' = \frac{328.2}{330/500} = \frac{328.2 \times 500}{330} = 497.27$$

$$\therefore \text{beats heard by B} = 500 - 497.27 = 2.73 \text{ Hz}$$

EXERCISES 26

1. If the velocity of sound in air at 15°C is 342 metres per second calculate the velocity at (a) 0°C , (b) 47°C . What is the velocity if the pressure of the air changes from 76 cm to 75 cm mercury, the temperature remaining constant at 15°C ?

2. Describe a determination (other than resonance) of the velocity of sound in air. How is the velocity dependent upon atmospheric conditions? Give Newton's expression for the velocity of sound in a gas, and Laplace's correction. Hence calculate the velocity of sound in air at 27°C . (Density of air at S.T.P. = 1.29 kg m^{-3} ; $C_p = 1.02\text{ kJ kg}^{-1}\text{ K}^{-1}$; $C_v = 0.72\text{ kJ kg}^{-1}\text{ K}^{-1}$.) (L.)

3. Describe the factors on which the velocity of sound in a gas depends. A man standing at one end of a closed corridor 57 m long blew a short blast on a whistle. He found that the time from the blast to the sixth echo was two seconds. If the temperature was 17°C , what was the velocity of sound at 0°C ? (C.)

4. Describe an experiment to find the velocity of sound in air at room temperature.

A ship at sea sends out simultaneously a wireless signal above the water and a sound signal through the water, the temperature of the water being 4°C . These signals are received by two stations, *A* and *B*, 40 km apart, the intervals between the arrival of the two signals being $16\frac{1}{2}\text{ s}$ at *A* and 22 s at *B*. Find the bearing from *A* of the ship relative to *AB*. The velocity of sound in water at $t^{\circ}\text{C}$ $\text{m s}^{-1} = 1427 + 3.3t$. (N.)

5. Write down an expression for the speed of sound in an ideal gas. Give a consistent set of units for the quantities involved.

Discuss the effect of changes of pressure and temperature on the speed of sound in air.

Describe an experimental method for finding a *reliable* value for the speed of sound in free air. (N.)

6. Describe an experiment to measure the velocity of sound in the open air. What factors may affect the value obtained and in what way may they do so?

It is noticed that a sharp tap made in front of a flight of stone steps gives rise to a ringing sound. Explain this and, assuming that each step is 0.25 m deep, estimate the frequency of the sound. (The velocity of sound may be taken to be 340 m s^{-1} .) (L.)

7. Explain why sounds are heard very clearly at great distances from the source (a) on still mornings after a clear night, and (b) when the wind is blowing from the source to the observer. (W.)

8. Describe one or two experiments to test each of the following statements: (a) If two notes are recognised by ear to be of the same pitch their sources are making the same number of vibrations per sec. (b) The musical interval between two notes is determined by the ratio of the frequencies of the vibrating sources of the notes. (L.)

9. Give a brief account of any important and characteristic wave phenomena which occur in sound. Why are sound waves in air regarded as longitudinal and not transverse?

An observer looking due north sees the flash of a gun 4 seconds before he records the arrival of the sound. If the temperature is 20°C and the wind is blowing from east to west with a velocity of 48 km per hour , calculate the distance between the observer and the gun. The velocity of sound in air at 0°C is 330 m s^{-1} . Why does the velocity of sound in air depend upon the temperature but not upon the pressure? (N.)

10. Explain upon what properties and conditions of a gas the velocity of sound through it depends.

Describe, and explain in detail, a laboratory method of measuring the velocity of sound in air. (L.)

Beats

11. Explain how *beats* are produced by two notes sounding together and obtain an expression for the number of beats heard per second.

A whistle of frequency 1000 Hz is sounded on a car travelling towards a cliff with a velocity of 18 m s^{-1} , normal to the cliff. Find the apparent frequency of the echo as heard by the car driver. Derive any relations used. (Assume velocity of sound in air to be 330 m s^{-1} .) (L.)

12. What is meant by (a) the *amplitude*, (b) the *frequency* of a vibration in the atmosphere? What are the corresponding characteristics of the musical sound associated with the vibration? How would you account for the difference in quality between two notes of the same pitch produced by two different instruments, e.g., by a violin and by an organ pipe?

What are 'beats'? Given a set of standard forks of frequencies 256, 264, 272, 280, and 288, and a tuning-fork whose frequency is known to be between 256 and 288, how would you determine its frequency to four significant figures? (W.)

13. Explain the origin of the beats heard when two tuning-forks of slightly different frequency are sounded together. Deduce the relation between the frequency of the beats and the difference in frequency of the forks. How would you determine which fork had the higher frequency?

A simple pendulum set up to swing in front of the 'seconds' pendulum ($T = 2 \text{ s}$) of a clock is seen to gain so that the two swing in phase at intervals of 21 s. What is the time of swing of the simple pendulum? (L.)

Doppler's Principle

14. An observer beside a railway line determines the speed of a train by observing the change in frequency of the note of its whistle as it passes him. Explain why a change of frequency occurs and derive the relation from which the speed may be calculated. Describe an example of the same principle in another branch of physics.

Find the lowest velocity that can be measured in this way, if the true frequency of the whistle is 1000 Hz and the observer is unable to detect departures from this frequency of less than 20 Hz. (Assume the velocity of sound to be 340 m s^{-1} .) (L.)

15. Explain what is meant by the *Doppler effect* in sound. Does an observer hear the same pitch from a given source of sound irrespective of whether the source approaches the stationary observer at a certain velocity or the observer approaches the stationary source at the same velocity? Explain how you arrived at your answer.

The light of the H (calcium) line of the spectrum is deviated through an angle of $45^\circ 12'$ by a certain prism. When observed in the light of a distant nebula, the deviation is $44^\circ 15'$. Calculate the velocity of the nebula in the line of sight, taking the velocity of light in vacuo to be $3.00 \times 10^8 \text{ m s}^{-1}$ and the deviation to be inversely proportional to the wavelength of the light over the range of values to be considered. (L.)

16. Explain in each case the change in the apparent frequency of a note

brought about by the motion of (i) the source, (ii) the observer, relative to the transmitting medium.

Derive expressions for the ratio of the apparent to the real frequency in the cases where (a) the source, (b) the observer, is at rest, while the other is moving along the line joining them.

The locomotive of a train approaching a tunnel in a cliff face at 95 km.p.h. is sounding a whistle of frequency 1000 Hz. What will be the apparent frequency of the echo from the cliff face heard by the driver? What would be the apparent frequency of the echo if the train were emerging from the tunnel at the same speed? (Take the velocity of sound in air as 330 m s^{-1} .) (L.)

17. (a) State the conditions necessary for 'beats' to be heard and derive an expression for their frequency.

(b) A fixed source generates sound waves which travel with a speed of 330 m s^{-1} . They are found by a distant stationary observer to have a frequency of 500 Hz. What is the wavelength of the waves? From first principles find (i) the wavelength of the waves in the direction of the observer, and (ii) the frequency of the sound heard if (1) the source is moving towards the stationary observer with a speed of 30 m s^{-1} , (2) the observer is moving towards the stationary source with a speed of 30 m s^{-1} , (3) both source and observer move with a speed of 30 m s^{-1} and approach one another. (N.)

18. What is the *Doppler effect*? Find an expression for it when the observer is at rest and there is no wind.

A whistle is whirled in a circle of 100 cm radius and traverses the circular path twice per second. An observer is situated outside the circle but in its plane. What is the musical interval between the highest and lowest pitch observed if the velocity of sound is 332 m s^{-1} ? (L.)

19. Explain why the frequency of a wave motion appears, to a stationary observer, to change as the component of the velocity of the source along the line joining the source and observer changes. Describe two illustrations of this effect, one with sound and one with light.

A stationary observer is standing at a distance l from a straight railway track and a train passes with uniform velocity v sounding a whistle with frequency n_0 . Taking the velocity of sound as V , derive a formula giving the observed frequency n as a function of the time. At which position of the train will $n = n_0$? Give a physical interpretation of the result. (C.)

Sound Intensity. Acoustics

20. Explain what is meant by (a) an intensity level in sound, (b) the statement that two intensity levels differ by 5 decibels. What considerations have determined the choice of a zero level in connection with the specification of loudness?

A loudspeaker produces a sound intensity level of 8 decibels above a certain reference level at a point P, 40 m from it. Find (a) the intensity level at a point 30 m from the loudspeaker, (b) the intensity level at P if the electrical power to the loudspeaker is halved. (L.)

21. Distinguish between the *intensity* and *loudness* of a sound. In what units would the intensity be measured? Define in each instance a unit employed to compare (a) the intensity and (b) the loudness of two sounds.

A source of sound is situated midway, between an observer and a flat wall. If the absorption coefficient of the wall is 0.25 find the ratio of the intensities of sound heard by the observer directly and by reflexion. Give the answer in decibels. (L.)

22. Describe a method for the accurate measurement of the velocity of sound in *free* air.

Indicate the factors which influence the velocity and how they are allowed for or eliminated in the experiment you describe.

At a point 20 m from a small source of sound the intensity is $0.5 \text{ microwatt cm}^{-2}$. Find a value for the rate of emission of sound energy from the source, and state the assumptions you make in your calculation. (N.)

23. Distinguish between *intensity* and *intensity level* of a sound.

The time taken for a sound to decay to one-millionth of its previous intensity after the source has been cut off is called the reverberation time. For a pure tone which gives an intensity level of 83 decibels in an empty lecture theatre the reverberation time was found to be 3.8 seconds. Calculate the sound intensity 7.6 seconds after the note was switched off. (Assume that the reference zero of intensity was $10^{-12} \text{ watt m}^{-2}$.)

Explain what was meant by a listener who stated that the note had a loudness of 70 phons.

Discuss how the acoustic properties of this lecture theatre might be improved. (N.)

24. (a) Discuss the relation between the *intensity level* and the *loudness* of a sound. Define suitable units in which each may be expressed.

(b) Give an account of the effect on the acoustics of a concert hall of such factors as: the design and material of the walls; the size of the audience; the frequency of the note. (L.)

25. A hall is 25 m long, 8 m wide and has walls 8 m high. The ceiling is a barrel vault of radius 5 m and the ends of the hall are plane. The floor is wood block and the walls are hard plaster, wood panelling and glass. The ceiling is also of hard plaster.

Indicate and give reasons for *three* defects of this hall as an auditorium and show how you would attempt to correct them.

Diagrams are essential in the answer to this question. (N.)

chapter twenty-seven

Vibrations in Pipes, Strings, Rods

Introduction

THE music from an organ, a violin, or a xylophone is due to vibrations in the air set up by oscillations in these instruments. In the organ, air is blown into a pipe, which sounds its characteristic note as the air inside it vibrates; in the violin, the strings are bowed so that they oscillate; and in a xylophone a row of metallic rods are struck in the middle with a hammer, which sets them into vibration.

Before considering each of the above cases in more detail, it would be best to consider the feature common to all of them. A violin string is fixed at both ends, A, B, and waves travel along m , n to each end of the string when it is bowed and are there reflected, Fig. 27.1 (i).

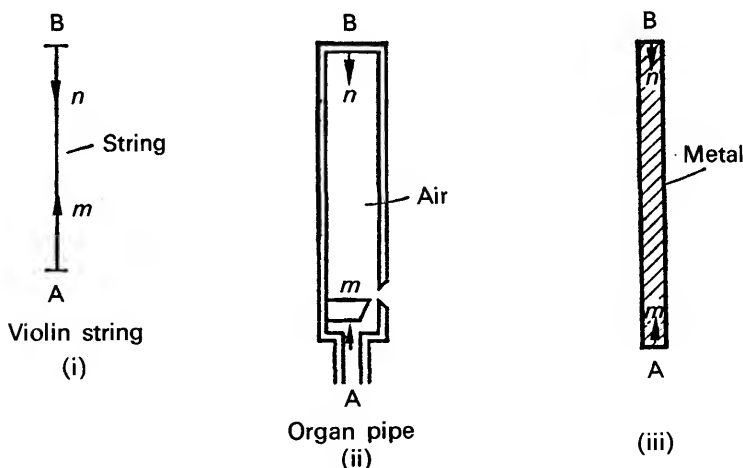


FIG. 27.1. Reflection of waves in instruments.

The vibrations of the particles of the string are hence due to *two waves of the same frequency and amplitude travelling in opposite directions*. A similar effect is obtained with an organ pipe closed at one end B, Fig. 27.1 (ii). If air is blown into the pipe at A, a wave travels along the direction m and is reflected at B in the opposite direction n . The vibrations of the air in the pipe are thus due to two waves travelling in opposite directions. If a metal rod is fixed at its middle in a vice and stroked at one end A, a wave travels along the rod in the direction m and is reflected at the other end B in the direction n , Fig. 27.1 (iii). The vibrations of the

rod, which produce a high-pitched note, are thus due to two waves travelling in opposite directions.

The resultant effect of two waves travelling in opposite directions with equal amplitude and frequency can easily be demonstrated. A light string, or thread, is tied to the end of a clapper, P, of an electric bell, and the other end of the string is passed round a grooved wheel, Fig. 27.2.

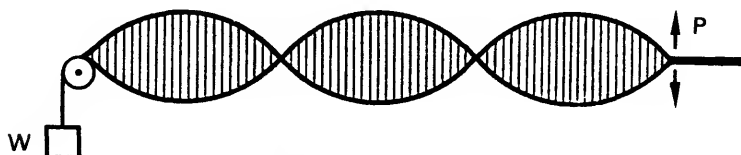


FIG. 27.2. Demonstration of stationary wave.

When the clapper vibrates, and a suitable weight W is attached to the string, a number of *stationary loops* is observed along the vibrating string, somewhat as shown in Fig. 27.2. By altering W a different number of stationary loops can be obtained. The wave along the string is known as a *stationary wave*, and we shall now discuss the formation of a stationary wave in detail.

Stationary Waves

Consider a plane-progressive wave a travelling in air along OA , Fig. 27.3. If it meets a wall at W a reflected wave b is obtained, and the condition of the air along W is due to the combined effects of a , b .

The layer of air at W must always be at rest since it is in contact with a fixed wall. For convenience, suppose that the displacements of the layers of air due to a at the instant shown are those represented by the sine wave in Fig. 27.3 (i), so that the displacement of the layer at W due to the incident wave is a maximum. Since the layer at W is always at rest, the displacements of the layers due to the wave *reflected* from the wall must be represented by the curve b at the same instant; otherwise the net displacement at W , which is the algebraic sum of WR , WH , will not be zero. From the curves a , b shown in Fig. 27.3 (i), it follows that the wave b reflected by the wall is 180° out of phase with the incident wave a .

At the instant t represented in Fig. 27.3 (i), the algebraic sum, S , of the displacements of the layers *everywhere* along OW is zero if the amplitudes of a , b are equal and the curves have the same wavelength. At an instant $T/4$ later, where T is the period of vibration of the layers the displacements of the layer due to the incident and reflected waves are those shown in Fig. 27.3 (ii). This can best be understood by imagining the incident wave a to have advanced to the right by $\frac{1}{4}$ -wavelength, and the reflected wave to have advanced to the left by $\frac{1}{4}$ -wavelength, which implies that the vibrating layers have now reached a displacement corresponding to a time $T/4$ later than t . The algebraic sums S of the displacement is then represented by the curve S in Fig. 27.3 (ii). At the end of a further time $T/4$, the displacements due to a , b are those shown

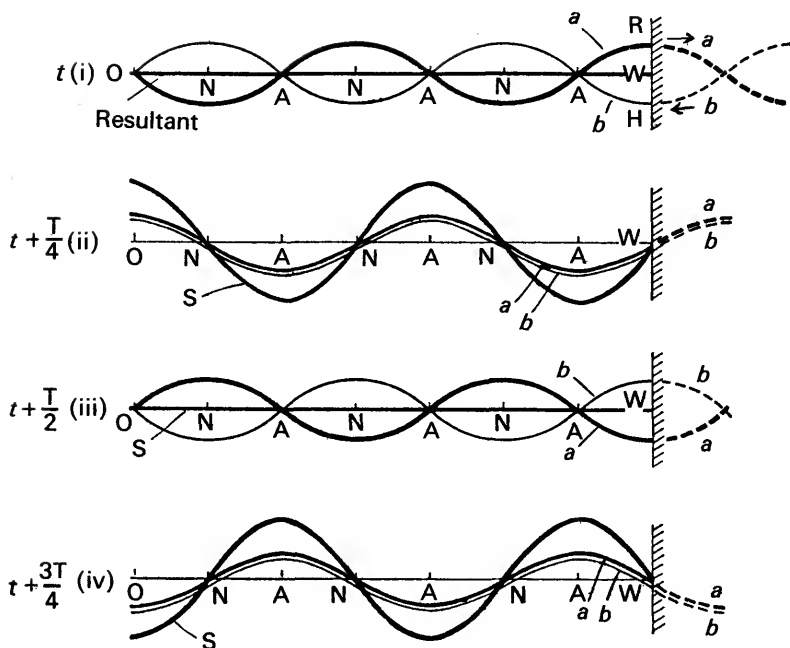


FIG. 27.3. Formation of stationary waves.

in Fig. 27.3 (iii); the waves have now advanced another $\frac{1}{4}$ -wavelength in opposite directions. The algebraic sum S of the displacements is again zero everywhere along OW at this instant. After a further time $T/4$, the displacements of the layers, and the resultant displacement S , are those shown in Fig. 27.3 (iv). The wave in the air represented by S is called a stationary wave.

Nodes and Antinodes

We have now sufficient information to deduce the conditions of the layers of air along OW when a stationary wave is obtained. From the curves showing the resultant displacement, S , in Fig. 27.3, it can be seen that some layers, marked N , are *permanently* at rest; these are known as **nodes**. The layers marked A , however, are vibrating through an amplitude twice as big as the incident or reflected waves, see Fig. 27.3 (ii) and (iv), and these are known as **antinodes**. Layers between consecutive nodes are vibrating in phase with each other, but the amplitude of vibration varies from zero at a node N to a maximum

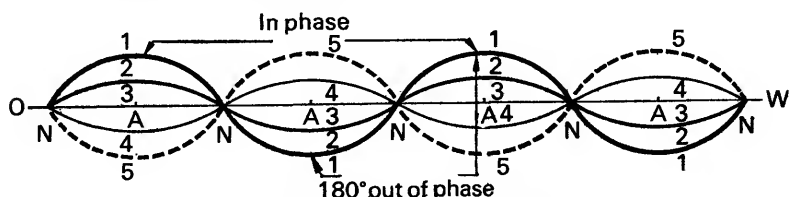


FIG. 27.4. Nodes and antinodes.

at the antinode A. Fig. 27.4 represents the displacement of the layers along OW at five different instants 1, 2, 3, 4, 5. It follows that

$$\text{the distance between consecutive nodes, } NN = \frac{\lambda}{2} \quad . \quad . \quad . \quad (i)$$

where λ is the wavelength of the stationary wave;

$$\text{the distance between consecutive antinodes, } AA = \frac{\lambda}{2}, \quad . \quad . \quad . \quad (ii)$$

and

$$\text{the distance from a node to the next antinode, } NA = \frac{\lambda}{4} \quad . \quad . \quad . \quad (iii)$$

The importance of the nodes and antinodes in a stationary wave lies in their simple connection with the wavelength.

Differences Between Plane-Progressive and Stationary Waves

At the beginning of Chapter 25, we considered in detail the plane-progressive wave and its effect on the medium (pp. 584 and 587). It was then shown that each layer vibrates with constant amplitude at the same frequency, and that each layer is out of phase with others near to it. When a stationary wave is present in a medium, however, some layers (nodes) are permanently at rest; others between the nodes are vibrating in phase with different amplitudes, increasing to a maximum at the antinodes. A stationary wave is always set up when two plane-progressive waves of equal amplitude and frequency travel in opposite directions in the same medium.

Mathematical proof of stationary wave properties. The properties of the stationary wave, already deduced, can be obtained easily from a mathematical

treatment. Suppose $y_1 = a \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda} \right)$ is a plane-progressive wave travel-

ling in one direction along the x -axis (p. 587). Then $y_2 = a \sin 2\pi \left(\frac{t}{T} + \frac{x}{\lambda} \right)$ represents a wave of the same amplitude and frequency travelling in the opposite direction. The resultant displacement, y , is hence given by

$$y = y_1 + y_2 = a \left[\sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda} \right) + \sin 2\pi \left(\frac{t}{T} + \frac{x}{\lambda} \right) \right]$$

$$\text{from which} \quad y = 2a \sin \frac{2\pi t}{T} \cdot \cos \frac{2\pi x}{\lambda} \quad . \quad . \quad . \quad . \quad . \quad . \quad (i)$$

using the transformation of the sum of two sine functions to a product.

$$\therefore y = B \sin \frac{2\pi t}{T} \quad . \quad . \quad . \quad . \quad . \quad . \quad (ii)$$

$$\text{where} \quad B = 2a \cos \frac{2\pi x}{\lambda} \quad . \quad . \quad . \quad . \quad . \quad . \quad (iii)$$

From (ii), B is the magnitude of the *amplitude* of vibration of the various layers; and from (iii) it also follows that the amplitude is a maximum and equal to $2a$ at $x = 0$, $x = \lambda/2$, $x = \lambda$, and so on. These points are thus antinodes, and consecutive antinodes are hence separated by a distance $\lambda/2$. The amplitude B is zero when $x = \lambda/4$, $x = 3\lambda/4$, $x = 5\lambda/4$, and so on. These points are thus nodes, and they are hence midway between consecutive antinodes.

The particle velocity in a stationary wave is the rate of change of the displacement (y) of the particle with respect to time (t). The velocity at the nodes is always zero since the particles there are permanently at rest. The velocity at an antinode increases from zero (when the particle is at the end of its oscillation) to a maximum (when the particle passes through its mean or original position). The corresponding displacement and velocity curves in the latter case are illustrated in Fig. 27.5 by P, M

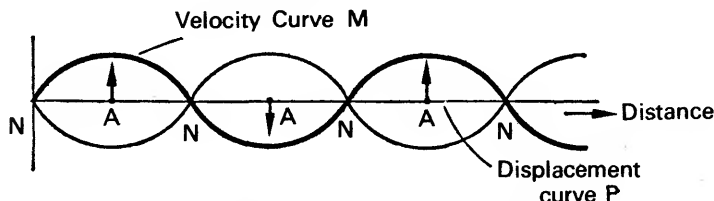


FIG. 27.5. Particle velocity due to stationary wave.

respectively. It will be noted that particles at neighbouring antinodes are moving in opposite directions at any instant.

Variation of Pressure in the Stationary Wave

Having considered the variation of the displacements and the particle velocities when a stationary wave travels in air, we must now turn our attention to the variation of *pressure* in the air.

Suppose that curve 1 represents the displacements at the antinodes and other points at an instant when they are a maximum, Fig. 27.6 (i). The layer of air immediately to the left of the node at a is then displaced towards a , since the displacement is positive from curve 1, and the layer immediately to the right of a also displaced towards a . The air at a is thus compressed, and the pressure is thus greater than normal, as represented by curve 1 in Fig. 27.6 (ii). The displacements of the layers on

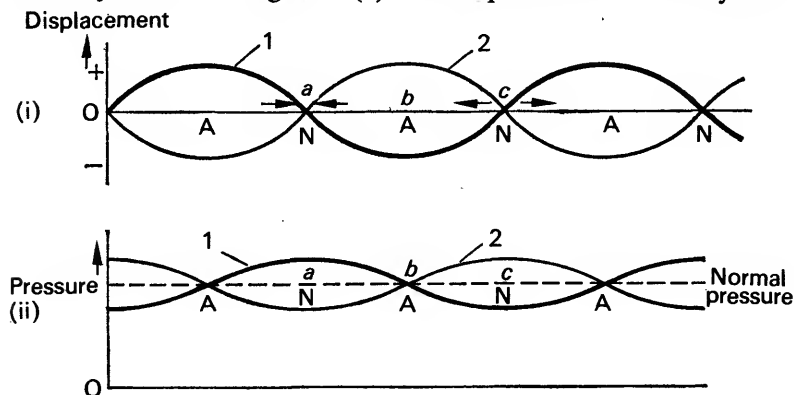


FIG. 27.6. Pressure variation due to stationary wave.

either side of the antinode at b are each a maximum to the left, and hence the pressure of the air is normal. The air on the left of the node c is

displaced away from a , and the air on the right of c is also displaced away from c . The air is thus rarified here, and hence the pressure is less normal. By carrying out the same procedure at other points in the air, it can be seen that the pressure variation corresponds to the curve 1 in Fig. 27.6 (ii).

When the displacements change to those represented by curve 2 in Fig. 27.6 (i), the variation of pressure at the same instant is shown by curve 2 in Fig. 27.6 (ii). We can now see that *the pressure variation is always a maximum at a node of the stationary wave, and is always zero at an antinode of the stationary wave*. In a plane-progressive wave, however, the pressure variation is the same at every point in a medium (p. 585).

EXAMPLE

Distinguish between progressive and stationary wave motion. Describe and illustrate with an example how stationary wave motion is produced. Plane sound waves of frequency 100 Hz fall normally on a smooth wall. At what distances from the wall will the air particles have (a) maximum, (b) minimum amplitude of vibration? Give reasons for your answer. (The velocity of sound in air may be taken as 340 m s^{-1} .) (L.)

First part. See p. 587 and p. 643.

Second part. A stationary wave is set up between the source and the wall, due to the production of a reflected wave. The wall is a node, since the air in contact with it cannot move; and other nodes are at equal distances, d , from the wall. But if the wavelength is λ ,

$$d = \frac{\lambda}{2} \text{ (p. 643).}$$

Also
$$\lambda = \frac{V}{f} = \frac{340}{100} = 3.4 \text{ m}$$

$$\therefore d = \frac{3.4}{2} = 1.7 \text{ m.}$$

Thus minimum amplitude of vibration is obtained 1.7, 3.4, 5.1 m . . . from the wall.

The antinodes are midway between the nodes. Thus maximum amplitude of vibration is obtained 0.85, 2.55, 4.25m, . . . from the wall.

VIBRATIONS OF AIR IN PIPES

Closed Pipe

A *closed* or *stopped organ pipe* consists essentially of a metal pipe closed at one end Q, and a blast of air is blown into it at the other end P, Fig. 27.7 (i). A wave thus travels up the pipe to Q, and is reflected at this end down the pipe, so that a *stationary wave* is obtained. The end Q of the closed pipe must be a node N, since the layer in contact with Q must be permanently at rest, and the open end A, where the air is free to vibrate, must be an antinode A. The simplest stationary wave in the

air in the pipe is hence represented by g in Fig. 27.7 (ii), where the pipe is positioned horizontally to show the relative displacement, y , of the layers at different distances, x , from the closed end Q; the axis of the stationary wave is Qx.

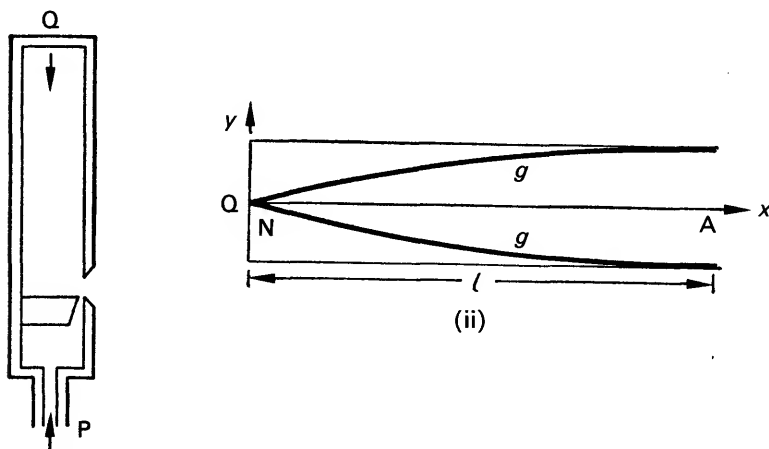


FIG. 27.7. (i). Closed (stopped) pipe. (ii). Fundamental of closed (stopped) pipe.

It can now be seen that the length l of the pipe is equal to the distance between a node N and a consecutive antinode A of the stationary wave. But $NA = \lambda/4$, where λ is the wavelength (p. 643).

$$\therefore \frac{\lambda}{4} = l$$

$$\therefore \lambda = 4l$$

But the frequency, f , of the note is given by $f = V/\lambda$, where V is the velocity of sound in air.

$$\therefore f = \frac{V}{4l}$$

This is the frequency of the lowest note obtainable from the pipe, and it is known as its *fundamental*. We shall denote the fundamental frequency by f_0 , so that

$$f_0 = \frac{V}{4l} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Overtones of Closed Pipe

If a stronger blast of air is blown into the pipes, notes of higher frequency can be obtained which are simple multiples of the fundamental frequency f_0 . Two possible cases of stationary waves are shown in Fig. 27.8. In each, the closed end of the pipe is a node, and the open end is an antinode. In Fig. 27.8 (i), however, the length l of the pipe is

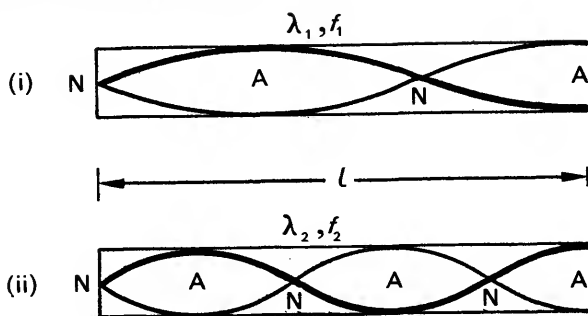


FIG. 27.8. Overtones in closed pipe.

related to the wavelength λ_1 of the wave by

$$l = \frac{3}{4} \lambda_1$$

$$\therefore \lambda_1 = \frac{4l}{3}$$

The frequency f_1 of the note is thus given by

$$f_1 = \frac{V}{\lambda_1} = \frac{3V}{4l} \quad \dots \quad (i)$$

But

$$f_0 = \frac{V}{4l}$$

$$\therefore f_1 = 3f_0 \quad \dots \quad (ii)$$

In Fig. 27.8 (ii), when a note of frequency f_2 is obtained, the length l of the pipe is related to the wavelength λ_2 by

$$l = \frac{5\lambda_2}{4}$$

$$\therefore \lambda_2 = \frac{4l}{5}$$

$$\therefore f_2 = \frac{V}{\lambda_2} = \frac{5V}{4l} \quad \dots \quad (iii)$$

$$\therefore f_2 = 5f_0 \quad \dots \quad (iv)$$

By drawing other sketches of stationary waves, with the closed end as a node and the open end as an antinode, it can be shown that higher frequencies can be obtained which have frequencies of $7f_0$, $9f_0$, and so on. They are produced by blowing harder at the open end of the pipe. The frequencies obtainable at a closed pipe are hence f_0 , $3f_0$, $5f_0$, and so on, i.e., the closed pipe gives only odd harmonics, and hence the frequencies $3f_0$, $5f_0$, etc. are possible *overtones*.

Open Pipe

An "open" pipe is one which is open at both ends. When air is blown into it at P, a wave m travels to the open end Q, where it is reflected in the direction n on encountering the free air, Fig. 27.9 (i). A stationary wave is hence set up in the air in the pipe, and as the two ends of the

pipe are open, they must both be *antinodes*. The simplest type of wave is hence that shown in Fig. 27.9 (ii), the x -axis of the wave being drawn along the middle of the pipe, which is horizontal. A node N is midway between the two antinodes.

The length l of the pipe is the distance between consecutive antinodes. But the distance between consecutive antinodes $= \lambda/2$, where λ is the wavelength (p. 643).

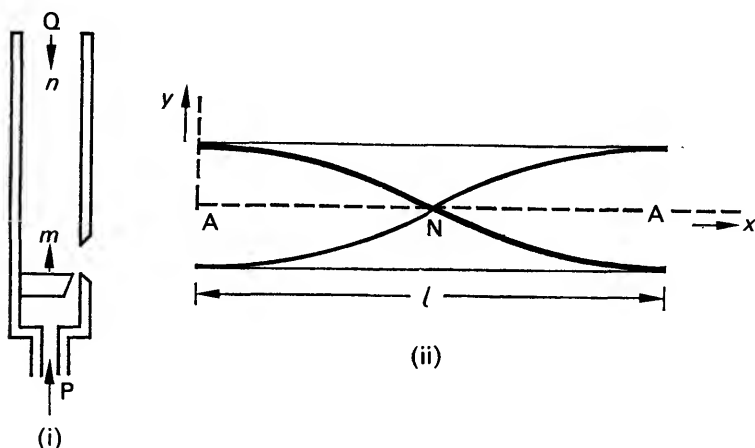


FIG. 27.9. (i). Open pipe. (ii). Fundamental of open pipe.

$$\therefore \frac{\lambda}{2} = l$$

$$\therefore \lambda = 2l$$

Thus the frequency f_0 of the note obtained from the pipe is given by

$$f_0 = \frac{V}{\lambda} = \frac{V}{2l} \quad \dots \quad (2)$$

This is the frequency of the fundamental note of the pipe.

Overtones of Open Pipe

Notes of higher frequencies than f_0 can be obtained from the pipe by blowing harder. The stationary wave in the pipe has always an antinode A at each end, and Fig. 27.10 (i) represents the case of a note of a frequency f_1 .

The length l of the pipe is equal to the wavelength λ_1 of the wave in this case. Thus

$$f_1 = \frac{V}{\lambda_1} = \frac{V}{l}$$

But

$$f_0 = \frac{V}{2l}, \text{ from (2) above.}$$

$$\therefore f_1 = 2f_0 \quad \dots \quad (i)$$

In Fig. 27.10 (ii), the length $l = \frac{3}{2} \lambda_2$, where λ_2 is the wavelength in the pipe. The frequency f_2 is thus given by

$$f_2 = \frac{V}{\lambda_2} = \frac{3V}{2l}$$

as $\lambda_2 = \frac{2l}{3}$.

$$\therefore f_2 = 3f_0 \quad \text{. (ii)}$$

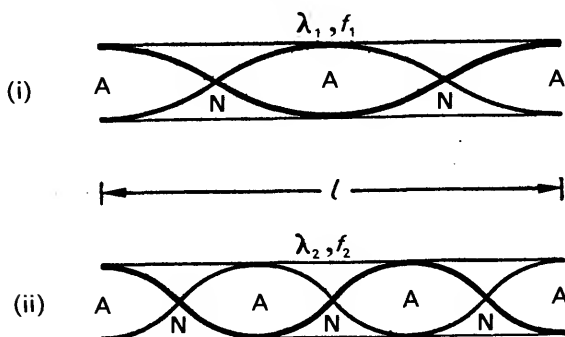


FIG. 27.10. Overtones of open pipes.

The frequencies of the overtones in the open pipe are thus $2f_0$, $3f_0$, $4f_0$, and so on, i.e., all harmonics are obtainable. The frequencies of the overtones in the closed pipe are $3f_0$, $5f_0$, $7f_0$, and so on, and hence the *quality* of the same note obtained from a closed and an open pipe is different (see p. 611).

Detection of Nodes and Antinodes, and Pressure Variation, in Pipes

The *nodes and antinodes* in a sounding pipe have been detected by suspending inside it a very thin piece of paper with lycopodium or fine sand particles on it, Fig. 27.11 (i). The particles are considerably agitated at the antinodes, but they are motionless at the nodes.

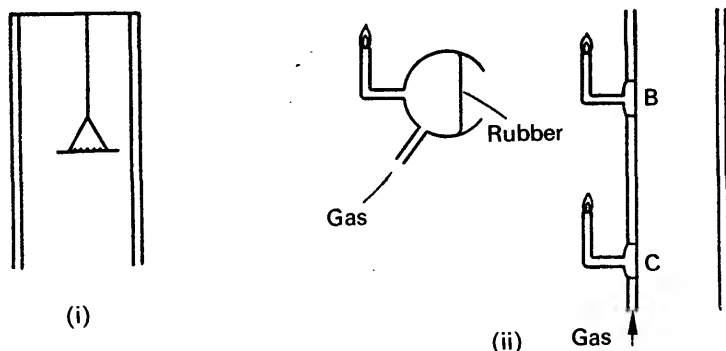


FIG. 27.11. (i). Detection of nodes and antinodes.
(ii). Detection of pressure.

The *pressure variation* in a sounding pipe has been examined by means of a sensitive flame, designed by Lord Rayleigh. The length of the flame can be made sensitive to the pressure of the gas supplied, so that if the pressure changes the length of flame is considerably affected. Several of the flames can be arranged at different parts of the pipe, with a thin rubber or mica diaphragm in the pipe, such as at B, C, Fig. 27.11 (ii). At a place of maximum pressure variation, which is a node (p. 645), the length of flame alters accordingly. At a place of constant (normal) pressure, which is an antinode, the length of flame remains constant.

The pressure variation at different parts of a sounding pipe can also be examined by using a suitable small microphone at B, C, instead of a flame. The microphone is coupled to a cathode-ray tube and a wave of maximum amplitude is shown on the screen when the pressure variation is a maximum. At a place of constant (normal) pressure, no wave is observed on the screen.

End-correction of Pipes

The air at the open end of a pipe is free to move, and hence the vibrations at this end of a sounding pipe extend a little into the air outside the pipe. The antinode of the stationary wave due to any note is thus a distance c from the open end in practice, known as the *end-correction*, and hence the wavelength λ in the case of a closed pipe is given by $\lambda/4 = l + c$, where l is the length of the pipe, Fig. 27.12 (i). In the case of an open pipe sounding its fundamental note, the wavelength λ is given by $\lambda/2 = l + c + c$, since *two* end-corrections are required, assuming the end-corrections are equal, Fig. 27.12 (ii). Thus $\lambda = 2(l + 2c)$. See also p. 648.

The mathematical theory of the end-correction was developed independently by Helmholtz and Rayleigh. It is now generally accepted

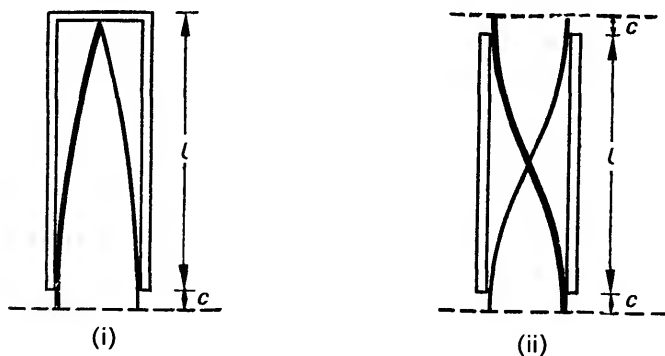


FIG. 27.12. (i). Closed pipe. (ii). Open pipe.

that $c = 0.58r$, or $0.6r$, where r is the radius of the pipe, so that the wider the pipe, the greater is the end-correction. It was also shown that the end-correction depends on the wavelength λ of the note, and tends to vanish for very short wavelengths.

Effect of Temperature, and End-correction, on the Pitch of Pipes

The frequency, f_0 , of the fundamental note of a closed pipe of length l and end-correction c is given by

$$f_0 = \frac{V}{\lambda} = \frac{V}{4(l+c)} \quad \text{. (i)}$$

with the usual notation, since $\lambda = 4(l+c)$. See p. 650. Now the velocity of sound, V , in air at $t^\circ\text{C}$ is related to its velocity V_0 at 0°C by

$$\frac{V}{V_0} = \sqrt{\frac{273+t}{273}} = \sqrt{1 + \frac{t}{273}} \quad \text{. (ii)}$$

since the velocity is proportional to the square root of the absolute temperature. Substituting for V in (i),

$$\therefore f_0 = \frac{V_0}{4(l+c)} \sqrt{1 + \frac{t}{273}} \quad \text{. (iii)}$$

From (iii), it follows that, with a given pipe, *the frequency of the fundamental increases as the temperature increases*. Also, for a given temperature and length of pipe, the frequency decreases as c increases. Now $c = 0.6r$, where r is the radius of the pipe. Thus *the frequency of the note from a pipe of given length is lower the wider the pipe*, the temperature being constant. The same results hold for an open pipe.

Resonance

If a diving springboard is bent and then allowed to vibrate freely, it oscillates with a frequency which is called its *natural frequency*. When a diver on the edge of the board begins to jump up and down repeatedly, the board is forced to vibrate at the frequency of the jumps; and at first, when the amplitude is small, the board is said to be undergoing *forced vibrations*. As the diver jumps up and down to gain increasing height for his dive, the frequency of the periodic downward force reaches a stage where it is practically the same as the natural frequency of the board. The amplitude of the board then becomes very large, and the periodic force is said to have set the board in *resonance*.

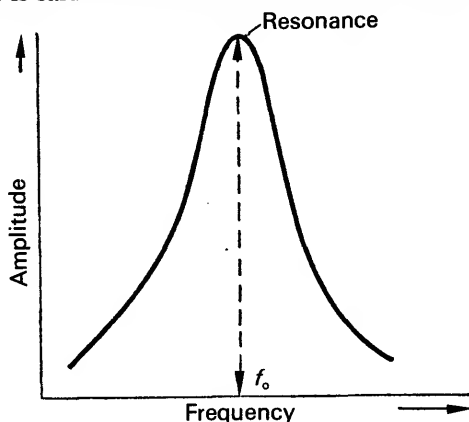


FIG. 27.13. Resonance curve.

A mechanical system which is free to move, like a wooden bridge or the air in pipes, has a natural frequency of vibration, f_0 , which depends on its dimensions. When a periodic force of a frequency different from f_0 is applied to the system, the latter vibrates with a small amplitude and undergoes forced vibrations. When the periodic force has a frequency equal to the natural frequency f_0 of the system, the amplitude of vibration becomes a maximum, and the system is then set into resonance. Fig. 27.13 is a typical curve showing the variation of amplitude with frequency. Some time ago it was reported in the newspapers that a soprano who was broadcasting had broken a glass tumbler on the table of a listener when she had reached a high note. This is an example of resonance. The glass had a natural frequency equal to that of the note sung, and was thus set into a vibration sufficiently violent to break it.

The phenomenon of resonance occurs in other branches of Physics than Sound and Mechanics. When an electrical circuit containing a coil and capacitor is "tuned" to receive the radio waves from a distant transmitter, the frequency of the radio waves is equal to the natural frequency of the circuit and resonance is therefore obtained. A large current then flows in the electrical circuit. A dark line in a continuous spectrum, an absorption line, is an example of *optical resonance*. Thus some of the yellow wavelengths from the sun's spectrum are absorbed by molecules of sodium vapour in the cooler part of the sun's atmosphere, which are set into resonance (see p. 464).

Sharpness of resonance. As the resonance condition is approached, the effect of the damping forces on the amplitude increases. Damping prevents the amplitude from becoming infinitely large at resonance. The lighter the damping, the sharper is the resonance, that is, the amplitude diminishes considerably at a frequency slightly different from the resonant frequency, Fig. 27.14. A heavily-damped system has a fairly flat resonance curve. Tuning is therefore more difficult in a system which has light damping.

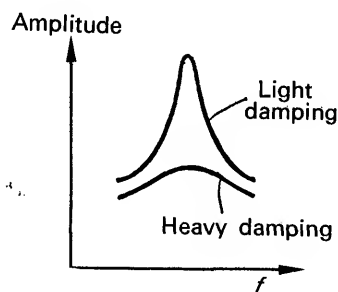


FIG. 27.14. Sharpness of resonance.

The effect of damping can be illustrated by attaching a simple pendulum carrying a pith bob, and one of the same length carrying a lead bob of equal size, to a horizontal string. The pendula are set into vibration by a third pendulum of equal length attached to the same string, and it is then seen that the amplitude of the lead bob is much greater than that of the pith bob. The damping of the pith bob due to air resistance is much greater than for the lead bob.

Resonance in a Tube or Pipe

If a person blows gently down a pipe closed at one end, the air inside vibrates freely, and a note is obtained from the pipe which is its funda-

mental (p. 646). A stationary wave then exists in the pipe, with a node N at the closed end and an antinode A at the open end, as explained on p. 645.



FIG. 27.15.
Resonance in
closed pipe.

If the prongs of a tuning-fork are held over the top of the pipe, the air inside it is set into vibration by the periodic force exerted on it by the prongs. In general, however, the vibrations are feeble, as they are *forced* vibrations, and the intensity of the sound heard is correspondingly small. But when a tuning-fork of the same frequency as the fundamental frequency of the pipe is held over the latter, the air inside is set in *resonance* by periodic force, and the amplitude of the vibrations is large. A loud note, which has the same frequency as the fork, is then heard coming from the pipe, and a stationary wave is set up with the top of the pipe acting as an antinode and the fixed end as a node, Fig. 27.15. If a sounding tuning-fork is held over a pipe open at both ends, resonance occurs when the stationary wave in the pipe has antinodes

at the two open ends, as shown by Fig. 27.9; the frequency of the fork is then equal to the frequency of the fundamental of the open pipe.

Resonance Tube Experiment. Measurement of Velocity of Sound and "End-Correction" of Tube

If a sounding tuning-fork is held over the open end of a tube T filled with water, resonance is obtained at some position as the level of water is gradually lowered, Fig. 27.16 (i). The stationary wave set up is then as

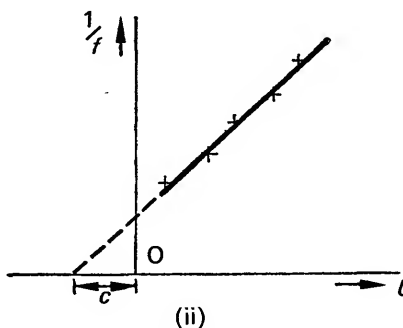
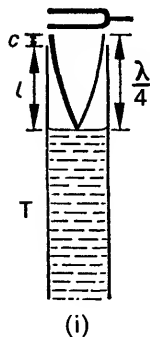


FIG. 27.16. Resonance tube experiment.

shown. If c is the end-correction of the tube (p. 650), and l is the length from the water level to the top of the tube, then

$$l + c = \frac{\lambda}{4} \quad . \quad . \quad . \quad (i)$$

But

$$\lambda = \frac{V}{f},$$

where f is the frequency of the fork and V is the velocity of sound in air.

$$\therefore l + c = \frac{V}{4f} \quad \dots \quad (ii)$$

If different tuning-forks of known frequency f are taken, and the corresponding values of l obtained when resonance occurs, it follows from equation (ii) that a graph of $1/f$ against l is a straight line, Fig. 27.16 (ii). Now from equation (ii), the gradient of the line is $4/V$; thus V can be determined. Also, the negative intercept of the line on the axis of l is c , from equation (ii); hence the end-correction can be found.

If only one fork is available, and the tube is sufficiently long, another method for V and c can be adopted. In this case the level of the water is lowered further from the position in Fig. 27.16 (i), until resonance is again obtained at a level L_1 , Fig. 27.17. Since the stationary wave set up is that shown and the new length to the top from L_1 is l_1 , it follows that

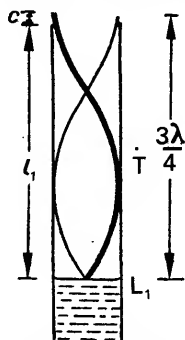


FIG. 27.17. Resonance at new water level.

$$l_1 + c = \frac{3\lambda}{4} \quad \dots \quad (iii)$$

But $l + c = \frac{\lambda}{4}$, from (ii).

Subtracting, $l_1 - l = \frac{\lambda}{2}$

$$\therefore \lambda = 2(l_1 - l)$$

$$\therefore V = f\lambda = 2f(l_1 - l) \quad \dots \quad (3)$$

In this method for V , therefore, the end-correction c is eliminated. The magnitude of c can be found from equations (ii) and (iii). Thus, from (ii),

$$3l + 3c = \frac{3\lambda}{4}$$

But, from (iii), $l_1 + c = \frac{3\lambda}{4}$

$$\therefore 3l + 3c = l_1 + c$$

$$\therefore 2c = l_1 - 3l$$

$$\therefore c = \frac{l_1 - 3l}{2} \quad \dots \quad (4)$$

Hence c can be found from measurements of l_1 and l .

EXAMPLES

1. Describe the natural modes of vibration of the air in an organ pipe closed at one end, and explain what is meant by the term "end-correction". A cylindrical pipe of length 28 cm closed at one end is found to be at resonance when a tuning fork of frequency 864 Hz is sounded near the open end.

Determine the mode of vibration of the air in the pipe, and deduce the value of the end-correction. [Take the velocity of sound in air as 340 m s^{-1} .] (*L.*)

First part. See text.

Second part. Let λ = the wavelength of the sound in the pipe.

Then
$$\lambda = \frac{V}{f} = \frac{34000}{864} = 39.35 \text{ cm}$$

If the pipe is resonating to its fundamental frequency f_0 , the stationary wave in the pipe is that shown in Fig. 27.16 and the wavelength λ_0 is given by $\lambda_0/4 = 28 \text{ cm}$. Thus $\lambda_0 = 112 \text{ cm}$. Since $\lambda = 39.35 \text{ cm}$, the pipe cannot be sounding its resonant frequency. The first overtone of the pipe is $3f_0$, which corresponds to a wavelength λ_1 given by $3\lambda/4 = 28$ (see Fig. 27.8).

$$\therefore \lambda_1 = \frac{112}{3} = 37\frac{1}{3} \text{ cm}$$

Consequently, allowing for the effect of an end-correction, the pipe is sounding its first overtone.

Let c = the end-correction in cm.

Then
$$28 + c = \frac{3\lambda_1}{4}$$

But, accurately,
$$\lambda_1 = \frac{V}{f} = \frac{34000}{864} = 39.35$$

$$\therefore 28 + c = \frac{3}{4} \times 39.35$$

$$\therefore c = 1.5 \text{ cm.}$$

2. Explain the phenomenon of resonance, and illustrate your answer by reference to the resonance-tube experiment. In such an experiment with a resonance tube the first two successive positions of resonance occurred when the lengths of the air columns were 15.4 cm and 48.6 cm respectively. If the velocity of sound in air at the time of the experiment was 34000 cm s^{-1} calculate the frequency of the source employed and the value of the end-correction for the resonance tube. If the air column is further increased in length, what will be the length when the next resonance occurs? (*W.*)

First part. See text.

Second part. Suppose c is the end-correction in cm. Then, from p. 654,

$$48.6 + c = \frac{3\lambda}{4} \quad . \quad . \quad . \quad (i)$$

and
$$15.4 + c = \frac{\lambda}{4} \quad . \quad . \quad . \quad (ii)$$

Subtracting
$$\therefore 33.2 = \frac{\lambda}{2}$$

$$\therefore 66.4 = \lambda \quad . \quad . \quad . \quad (iii)$$

$$\text{frequency, } f = \frac{V}{\lambda} = \frac{34000}{66.4} = 512 \text{ Hz}$$

The end-correction, c , is given by substituting $\lambda = 66.4$ in (i). Thus

$$48.6 + c = \frac{3}{4} \times 66.4$$

from which

$$c = 1.2 \text{ cm.}$$

velocity V of the transverse wave along it depends only on the values of T, s, l . The velocity is given by

$$V = \sqrt{\frac{T}{s/l}},$$

or $V = \sqrt{\frac{T}{m}} \quad \dots \dots \dots (5)$

where m is the "mass per unit length" of the string.

When T is in newtons and m in kilogramme per metre, then V is in metres per second.

The formula for V may be partly deduced by the method of dimensions, in which all the quantities concerned are reduced to the fundamental units of mass, M , length, L , and time, T . Suppose that

$$V = kT^x s^y l^z \quad \dots \dots \dots (i)$$

where k, x, y, z , are numbers. The dimensions of velocity V are LT^{-1} , the dimensions of tension T , a force, are MLT^{-2} , the dimension of s is M , and the dimension of l is L . As the dimensions on both sides of (i) must be equal, it follows that

$$LT^{-1} = (MLT^{-2})^x (M^y)(L^z)$$

Equating the indices of M, L, T on both sides, we have

$$\text{for } M, \quad x + y = 0$$

$$\text{for } L, \quad x + z = 1$$

$$\text{for } T, \quad 2x = 1$$

$$\therefore x = \frac{1}{2}, z = \frac{1}{2}, y = -\frac{1}{2}$$

Thus, from (i)

$$V = kT^{\frac{1}{2}} s^{-\frac{1}{2}} l^{\frac{1}{2}}$$

$$\therefore V = k \sqrt{\frac{Tl}{s}} = k \sqrt{\frac{T}{s/l}}$$

A rigid mathematical treatment shows that $k = 1$, since $V = \sqrt{\frac{T}{s/l}}$. Since s/l is the "mass per unit length" of the string, it follows that

$$V = \sqrt{\frac{T}{m}},$$

where m is the mass per unit length.

Modes of Vibration of Stretched String

If a wire is stretched between two points N, N and is plucked in the middle, a transverse wave travels along the wire and is reflected at the fixed end. A *stationary wave* is thus set up in the wire, and the simplest mode of vibration is one in which the fixed ends of the wire are nodes, N , and the middle is an antinode, A , Fig. 27.18. Since the distance be-

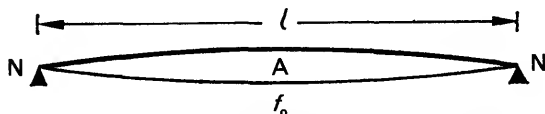


FIG. 27.18. Fundamental of stretched string.

tween consecutive nodes is $\lambda/2$, where λ is the wavelength of the transverse wave in the wire, it follows that

$$l = \frac{\lambda}{2},$$

where l is the length of the wire. Thus $\lambda = 2l$. The frequency f of the vibration is hence given by

$$f = \frac{V}{\lambda} = \frac{V}{2l},$$

where V is the velocity of the transverse wave. But $V = \sqrt{T/m}$, from previous.

$$\therefore f = \frac{1}{2l} \sqrt{\frac{T}{m}}$$

This is the frequency of the *fundamental* note obtained from the string; and if we denote the frequency by the usual symbol f_0 , we have

$$f_0 = \frac{1}{2l} \sqrt{\frac{T}{m}} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Overtones of Stretched String

The first overtone f_1 of a string plucked in the middle corresponds to a stationary wave shown in Fig. 27.19, which has nodes at the fixed ends and an antinode in the middle. If λ_1 is the wavelength, it can be seen that

$$l = \frac{3}{2} \lambda_1,$$

$$\text{or } \lambda_1 = \frac{2l}{3}.$$

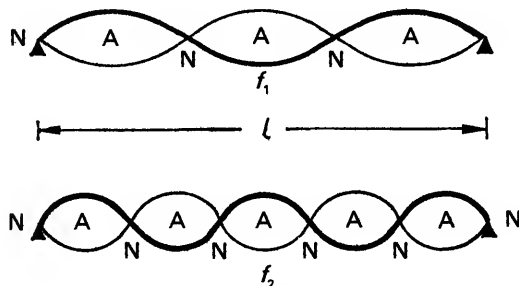


FIG. 27.19. Overtones of stretched string plucked in middle.

The frequency f_1 is thus given by

$$f_1 = \frac{V}{\lambda_1} = \frac{3V}{2l} = \frac{3}{2l} \sqrt{\frac{T}{m}} \quad . \quad . \quad . \quad . \quad . \quad (i)$$

But the fundamental frequency, $f_0 = \frac{1}{2l} \sqrt{\frac{T}{m}}$, from equation (6).

$$\therefore f_1 = 3f_0$$

The second overtone f_2 of the string when plucked in the middle

corresponds to a stationary wave shown in Fig. 27.19. In this case $l = \frac{5}{2}\lambda_2$, where λ_2 is the wavelength.

$$\therefore \lambda_2 = \frac{2l}{5}$$

$$\therefore f_2 = \frac{V}{\lambda_2} = \frac{5V}{2l}$$

where f_2 is the frequency. But $V = \sqrt{T/m}$.

$$\therefore f_2 = \frac{5}{2l} \sqrt{\frac{T}{m}} = 5f_0$$

The overtones are thus $3f_0$, $5f_0$, and so on.

Other notes than those considered above can be obtained by touching or "stopping" the string lightly at its midpoint, for example, so that the latter becomes a node in addition to those at the fixed ends. If the string is plucked one-quarter of the way along it from a fixed end, the simplest stationary wave set up is that illustrated in Fig. 27.20 (i). Thus the wavelength $\lambda = l$, and hence the frequency f is given by

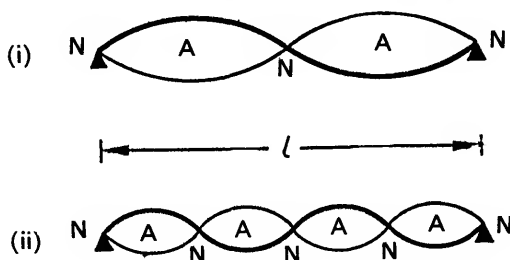


FIG. 27.20. Even harmonics in stretched string.

$$f = \frac{V}{\lambda} = \frac{V}{l} = \frac{1}{l} \sqrt{\frac{T}{m}}$$

$$\therefore f = 2f_0, \text{ since } f_0 = \frac{1}{2l} \sqrt{\frac{T}{m}}.$$

If the string is plucked one-eighth of the way from a fixed end, a stationary wave similar to that in Fig. 27.20 (ii) may be set up. The wavelength, $\lambda' = l/2$, and hence the frequency $f' = \frac{V}{\lambda'} = \frac{2V}{l}$.

$$\therefore f' = \frac{2}{l} \sqrt{\frac{T}{m}} = 4f_0$$

Verification of the Laws of Vibration of a Fixed String.

The Sonometer

As we have already shown (p. 658), the frequency of the fundamental of a stretched string is given by $f = \frac{1}{2l} \sqrt{\frac{T}{m}}$, writing f for f_0 . It thus,

follows that:

(1) $f \propto \frac{1}{l}$ for a given tension (T) and string (m constant).

(2) $f \propto \sqrt{T}$ for a given length (l) and string (m constant).

(3) $f \propto \frac{1}{\sqrt{m}}$ for a given length (l) and tension (T).

These are known as the "laws of vibration of a fixed string", first completely given by MERSENNE in 1636, and the *sonometer*, or *monochord*, was designed to verify them.

The sonometer consists of a hollow wooden box Q , with a thin horizontal wire attached to A on the top of it, Fig. 27.21. The wire passes

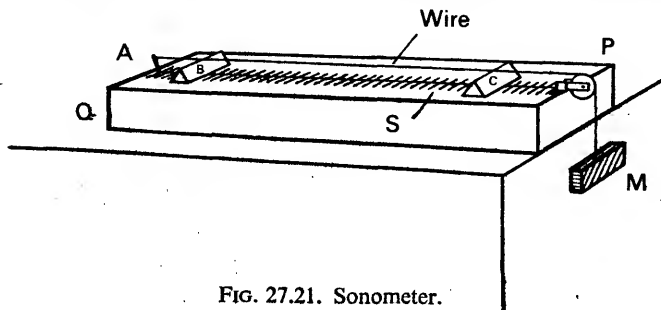


FIG. 27.21. Sonometer.

over a grooved wheel P , and is kept taut by a mass M hanging down at the other end. Wooden bridges, B , C , can be placed beneath the wire so that a definite length of wire is obtained, and the length of wire can be varied by moving one of the bridges. The length of wire between B , C can be read from a fixed horizontal scale S , graduated in centimetres, on the box below the wire.

(1) To verify $f \propto 1/l$ for a given tension (T) and mass per unit length (m), the mass M is kept constant so that the tension, T , in the wire is constant. The length, l , of the wire between B , C is varied by moving C until the note obtained by plucking BC in the middle is the same as that produced by a sounding tuning-fork of known frequency f . If the observer lacks a musical ear, the "tuning" can be recognised by listening for beats when the wire and the tuning-fork are both sounding, as in this case the frequencies of the two notes are nearly equal (p. 620). Alternatively, a small piece of paper in the form of an inverted V can be placed on the middle of the wire, and the end of the sounding tuning-fork then placed on the sonometer box. The vibrations of the fork are then transmitted through the box to the wire, which vibrates in resonance with the fork if its length is "tuned" to the note. The paper will then vibrate considerably and may be thrown off the wire.

Different tuning-forks of known frequency f are taken, and the lengths, l , of the wire are observed when they are tuned to the corresponding note. A graph of f against $1/l$ is then plotted, and is found to be a straight line within the limits of experimental error. Thus $f \propto 1/l$ for a given tension and mass per unit length of wire.

(2) To verify $f \propto \sqrt{T}$ for a given length and mass per unit length, the length BC between the bridges is kept fixed, so that the length of wire is constant, and the mass M is varied to alter the tension. The experimental difficulty to be overcome is how to find the frequency f of the note produced when the wire between B, C is plucked in the middle. For this purpose a second wire, fixed to R, S on the sonometer, is utilised, usually with a weight (not shown) attached to one end to keep the tension constant, Fig. 27.22. This wire has bridges P, N beneath it, and N is moved until the note from the wire between P, N is the same as the note from the wire between B, C. Now the tension in PN is constant as the wire is fixed to R, S. Thus, since frequency, $f, \propto 1/l$ for a given tension and wire, the frequency of the note from BC is proportional to $1/l$, where l is the length of PN.

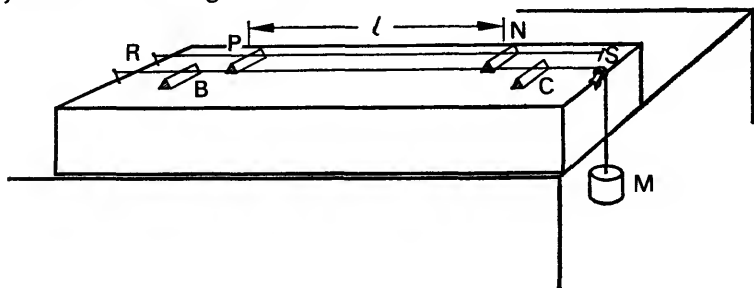


FIG. 27.22. Verification of $f \propto \sqrt{T}$.

If a different mass is attached to the end of the wire BC, the tension in the wire is altered. Keeping BC fixed, the bridge N is moved until the note from PN is the same as that obtained from BC, and the length PN (l) is again noted. Then, as before, the frequency of the new note in BC is proportional to $1/l$. By altering the mass M , and observing the corresponding length l , a graph of $1/l$ can be plotted against \sqrt{T} , where T is the weight of M . The graph is a straight line, within the limits of experimental error, and hence $1/l \propto \sqrt{T}$. Thus $f \propto \sqrt{T}$ for a given length of wire and mass per unit length.

(3) To verify $f \propto 1/\sqrt{m}$ for a given length and tension, wires of different material are connected to B, C, and the same mass M and the same length BC are taken. The frequency, f , of the note obtained from BC is again found by using the second wire RS in the way already described. The mass per unit length, m , is the mass per metre length of wire, and is given by $\pi r^2 \rho$ kg m⁻¹, where r is the radius of the wire in m and ρ is its density in kg m⁻³, as $(\pi r^2 \times 1)$ m³ is the volume of 1 m of the wire. Since $f \propto 1/l$, where l is the length on the second wire, a graph of $1/l$ against $1/\sqrt{m}$ should be a straight line; thus a graph of l against \sqrt{m} should be a straight line if $f \propto 1/\sqrt{m}$ for a given length and mass per unit length. Experiment shows this is the case.

Melde's Experiment on Vibrations in Stretched String

MELDE gave a striking demonstration of the stationary wave set up in a vibrating string. He used a light thread with one end attached to

the prong P of an electrically-maintained tuning-fork, and the other end connected to a weight W after passing over a grooved wheel Q, Fig. 27.23. With the vibrations of the prong perpendicular to the length

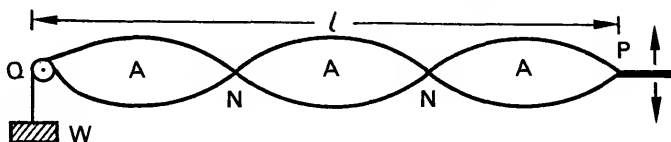


FIG. 27.23. Melde's experiment.

of the thread a number of "loops" can be observed, which are due to the rapid movement of the thread as the stationary transverse wave passes along it. The nodes of the stationary wave are at Q and at P, as the amplitude of vibration of the prong is small, and the number n of the loops depends on the frequency f of the fork, the length l of PQ, the tension T in the thread, and its mass per unit length, m .

The length of a loop = l/n . But this is the distance between consecutive nodes, which is $\lambda/2$, where λ is the wavelength of the stationary wave in the thread.

$$\begin{aligned}\therefore \frac{\lambda}{2} &= \frac{l}{n} \\ \therefore \lambda &= \frac{2l}{n} \\ \therefore f &= \frac{V}{\lambda} = \frac{n}{2l} V,\end{aligned}$$

where V is the velocity of the transverse wave. The frequency of the transverse wave is the same as that of the tuning-fork.

But
$$V = \sqrt{\frac{T}{m}} \text{ (p. 657)}$$

$$\therefore f = \frac{n}{2l} \sqrt{\frac{T}{m}}, \quad \dots \dots \dots (7)$$

$$\therefore n\sqrt{T} = \text{constant},$$

if f , T , m are kept constant. In this case, therefore,

$$n = \frac{\text{constant}}{\sqrt{T}}$$

$$\text{or } n^2 \propto \frac{1}{T} \quad \dots \dots \dots (8)$$

This relation can be verified by varying the tension T in the thread, and obtaining a corresponding whole number, n , of loops. A graph of n^2 against $1/T$ is then plotted, and is a straight line passing through the origin.

When the prong P of the tuning-fork is vibrating in the same direction as the length of string, the latter moves from position 3 to position 2 as the prong moves from one end a of an oscillation to the other end b ,

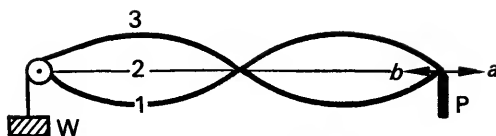


FIG. 27.24. Melde's experiment.

Fig. 27.24. As the string continues from position 2 to position 1, the prong moves back from b to a to complete 1 cycle of oscillation. The fork thus goes through one complete cycle in the same time as the particles of the string go through half a cycle, and hence the frequency of the transverse wave is *half* the frequency f of the fork, unlike the first case considered, Fig. 27.24. Instead of equation (7), we now have

$$\frac{f}{2} = \frac{n}{2l} \sqrt{\frac{T}{m}}$$

or $n = fl\sqrt{T/m},$

and hence with the same tension, thread, and length l , the number of loops n is half that obtained previously (p. 662).

Measurement of the Frequency of A.C. Mains

The frequency of the alternating current (A.C.) mains can be determined with the aid of a sonometer wire. The alternating current is passed into the wire MP, and the poles N, S of a powerful magnet are placed on either side of the wire so that the magnetic field due to it is perpendicular to the wire, Fig. 27.25. As a result of the magnetic effect of the current, a force acts on the wire which is perpendicular to the directions of both the magnetic field and the current, and hence the wire is subjected to a transverse force. If the current is an alternating one of 50 Hz, the magnitude of the force varies at the rate of 50 Hz. By adjusting the tension in the sonometer wire, whose magnitude is read on the spring-balance A, a position can be reached when the wire is seen to be vibrating through a large amplitude; in this case the wire is *resonating* to the applied force, Fig. 27.25.

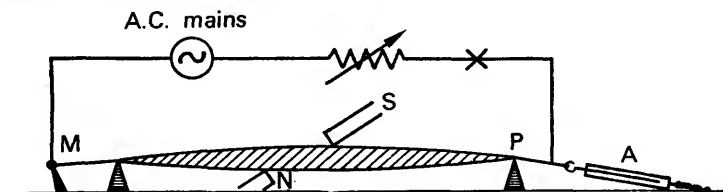


FIG. 27.25. Measurement of frequency of A.C. mains.

The length l of wire between the bridges is now measured, and the tension T and the mass per unit length, m , are also found. The frequency f of the alternating current is then given by

$$f = \frac{1}{2l} \sqrt{\frac{T}{m}}.$$

Velocity of Longitudinal Waves in Wires

If a sonometer wire is stroked along its length by a rosined cloth, a high-pitched note is obtained. This note is due to *longitudinal* vibrations in the wire, and must be clearly distinguished from the note produced when the wire is plucked, which sets up *transverse* vibrations of the wire and a corresponding transverse wave. As we saw on p. 594, the velocity V of a longitudinal wave in a medium is

$$V = \sqrt{\frac{E}{\rho}},$$

where E is Young's modulus for the wire and ρ is its density. The wavelength, λ , of the longitudinal wave is $2l$, where l is the length of the wire, since a stationary longitudinal wave is set up. Thus the frequency f of the note is given by

$$f = \frac{V}{\lambda} = \frac{1}{2l} \sqrt{\frac{E}{\rho}}.$$

The frequency of the note may be obtained with the aid of an audio oscillator, and thus the velocity of sound in the wire, or its Young's modulus, can be calculated.

EXAMPLES

1. Explain the meaning of the term *resonance*, giving in illustration two methods of obtaining resonance between the stretched string of a sonometer and a tuning-fork of fixed frequency. A sonometer wire of length 76 cm is maintained under a tension of value 4 kgf and an alternating current is passed through the wire. A horse-shoe magnet is placed with its poles above and below the wire at its midpoint, and the resulting forces set the wire in resonant vibration. If the density of the material of the wire is 8800 kg m^{-3} and the diameter of the wire is 1 mm, what is the frequency of the alternating current? (L .)

First part. See text.

Second part. The wire is set into resonant vibration when the frequency of the alternating current is equal to its natural frequency, f .

Now
$$f = \frac{1}{2l} \sqrt{\frac{T}{m}} \quad \dots \dots \dots (i)$$

where $l = 0.76 \text{ m}$, $T = 4 \times 9.8 \text{ newton}$, and $m = \text{mass per metre in kg m}^{-1}$.

Also,
$$\begin{aligned} \text{mass of 1 metre} &= \text{volume} \times \text{density} \\ &= \pi r^2 \times 1 \times 8800 \text{ kg,} \end{aligned}$$

where radius r of wire $= \frac{1}{2} \text{ mm} = 0.5 \times 10^{-3} \text{ m}$

From (i),
$$\therefore f = \frac{1}{2 \times 0.76} \sqrt{\frac{4 \times 9.8}{\pi \times 0.5^2 \times 10^{-6} \times 1 \times 8800}}$$
$$= 49.6 \text{ Hz.}$$

2. What data would be required in order to predict the frequency of the note emitted by a stretched wire (a) when it is plucked, (b) when it is stroked along its length? A weight is hung on the wire of a vertical sonometer. When the vibrating length of the wire is adjusted to 80 cm the note it emits when

plucked is in tune with a standard fork. On adding a further weight of 100 g the vibrating length has to be altered by 1 cm in order to restore the tuning. What is the initial weight on the wire? (*L*.)

First part. When the wire is plucked the vibrations of the particles produce transverse waves, and the frequency of the note is given by $f = \frac{1}{2l} \sqrt{\frac{T}{m}}$. When the wire is stroked along its length, the vibrations of the particles produce a *longitudinal* wave, and the velocity of the wave is given by $V = \sqrt{E/\rho}$, where E is Young's modulus for the wire and ρ is its density. The frequency in the latter case thus depends on the magnitudes of E and ρ , as well as on the length of the wire.

Second part. Let W = the initial weight on the wire in kgf,
and f = the frequency of the fork

Since
$$= \frac{1}{2l} \sqrt{\frac{T}{m}}$$

we have
$$f = \frac{1}{2 \times 0.80} \sqrt{\frac{Wg}{m}} \quad \dots \quad (i)$$

When a weight of 0.1 kgf is added, the frequency increases since the tension increases. The new length of the wire = 0.81 m.

$$\therefore f = \frac{1}{2 \times 0.81} \sqrt{\frac{(W + 0.1)g}{m}} \quad \dots \quad (ii)$$

From (i) and (ii), it follows that

$$\begin{aligned} \frac{1}{1.60} \sqrt{\frac{Wg}{m}} &= \frac{1}{1.62} \sqrt{\frac{(W + 0.1)g}{m}} \\ \therefore 162^2 W &= 160^2 (W + 0.1) \\ W &= \frac{160^2 \times 0.1}{162^2 - 160^2} = 4 \text{ kgf (approx.)} \end{aligned}$$

VIBRATIONS IN RODS

Sound waves travel through liquids and solids, as well as through gases, and in the nineteenth century an experiment to measure the velocity of sound in iron was carried out by tapping one end of a very long iron tube. The speed of sound in iron is much greater than in air, and the sound through the iron thus arrived at the other end of the pipe before the sound transmitted through the air. From a knowledge of the interval between the sounds, the length of the pipe, and the velocity of sound in air, the velocity of sound in iron was determined. More accurate methods were soon forthcoming for the velocity of sound in substances such as iron, wood, and glass, and they depend mainly on the formation of stationary waves in rods of these materials.

Consider a rod AA fixed by a vice B at its mid-point N, Fig. 27.26. If the rod is stroked along its length by a rosined cloth, a stationary longitudinal wave is set up in the rod on account of reflection at its ends, and a high-pitched note is obtained. Since the mid-point of the rod is fixed, this is a node, N, of the stationary wave; and since the ends

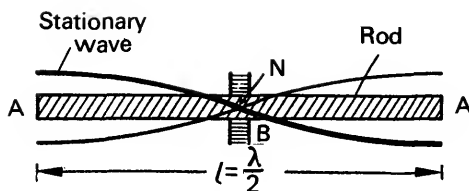


FIG. 27.26. Stationary wave in rod.

of the rod are free, these are antinodes, A. Thus the length l of the rod is equal to half the wavelength, $\lambda/2$, of the wave in the rod, and hence $\lambda = 2l$. Thus the velocity of the sound in the rod, $V = f\lambda = f \times 2l$, where f is the frequency of the note from the rod.

Kundt's Tube

About 1868, KUNDT devised a simple method of showing the stationary waves in air or any other gas. He used a closed tube T containing the gas, and sprinkled some dry lycopodium powder, or cork dust, along the entire length, Fig. 27.27. A rod AE, clamped at its mid-point, is placed with one end projecting into T, and a disc E is attached at this end so that it just clears the sides of the tube, Fig. 27.27. When the

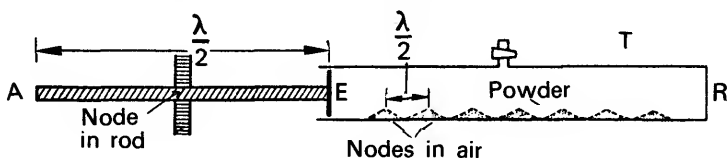


FIG. 27.27. Kundt's tube.

rod is stroked at A by a rosined cloth in the direction EA, the rod vibrates longitudinally and a high-pitched note can be heard. The end E acts as a vibrating source of the same frequency, and a sound wave thus travels through the air in T and is reflected at the fixed end R. If the rod is moved so that the position of E alters, a position can be found when the stationary wave in the air in T causes the lycopodium powder to become violently agitated. The powder then settles into definite small heaps at the nodes, which are the positions of permanent rest of the stationary wave, and the distance between consecutive nodes can best be found by measuring the distance between several of them and dividing by the appropriate number.

Determination of Velocity of Sound in a Rod

Kundt's tube can be used to determine the velocity of sound, V_r , in the rod. Suppose the length of the rod is l ; then $\lambda/2 = l$, or $\lambda = 2l$, where λ is the wavelength of the sound wave in the rod (p. 643). Thus the frequency of the high-pitched note obtained from the rod is given by

$$f = \frac{V_r}{\lambda} = \frac{V_r}{2l} \quad \dots \quad (i)$$

If l_1 is the distance between consecutive nodes of the stationary wave in the air, we have $\lambda_1/2 = l_1$, where λ_1 is the wavelength of the sound

wave in the air. Thus $\lambda_1 = 2l_1$, and hence the frequency of the wave, which is also f , is given by

$$f = \frac{V_a}{\lambda} = \frac{V_a}{2l_1}, \quad \dots \quad (ii)$$

where V_a is the velocity of sound in air. From (i) and (ii) it follows that

$$\frac{V_r}{2l} = \frac{V_a}{2l_1}$$

$$\therefore V_r = \frac{l}{l_1} V_a \quad \dots \quad (9)$$

Thus knowing V_a , l , l_1 , the velocity of sound in the rod, V_r , can be calculated. By using glass, brass, copper, steel and other substances in the form of a rod, the velocity of sound in these media have been determined. Kundt also used liquids in the tube T instead of air, and employed fine iron filings instead of lycopodium powder to detect the nodes in the liquid. In this way he determined the velocity of sound in liquids.

Determination of Young's Modulus of a Rod

On p. 622, it was shown that the velocity of sound, V , in a medium is always given by

$$V = \sqrt{\frac{E}{\rho}},$$

where E is the appropriate modulus of elasticity of the medium and ρ is its density. In the case of a rod undergoing longitudinal vibrations, as in Kundt's tube experiment, E is Young's modulus (see p. 594). Thus if V_r is the velocity of sound in the rod,

$$V_r = \sqrt{\frac{E}{\rho}},$$

and

$$\therefore E = V_r^2 \rho \quad \dots \quad (10)$$

Since V_r is obtained by the method explained above, and ρ can be obtained from tables, it follows that E can be calculated.

Determination of Velocity of Sound in a Gas

If the air in Kundt's tube T is replaced by some other gas, and the rod stroked, the average distance l' between the piles of dust in T is the distance between consecutive nodes of the stationary wave in the gas. The wavelength, λ_g , in the gas is thus $2l'$, and the frequency f is given by

$$f = \frac{V_g}{\lambda_g} = \frac{V_g}{2l'},$$

where V_g is the velocity of sound in the gas. But the wavelength, λ , of the wave in the rod $= 2l$, where l is the length of the rod (p. 666); hence f is also given by

$$f = \frac{V_r}{\lambda} = \frac{V_r}{2l}$$

The frequency f of the sound wave in A, B is the same, being the frequency of the note obtained from R. Since $f = V/\lambda$, it follows that

$$\frac{V_g}{\lambda_g} = \frac{V_a}{\lambda_a}, \quad \dots \quad (i)$$

where V_g , V_a are the velocities of sound in carbon dioxide and air respectively, and λ_g , λ_a are the corresponding wavelengths.

Now
$$\frac{\lambda_g}{\lambda_a} = \frac{d_b}{d_a} \quad \dots \quad (ii)$$

since the distance between successive nodes is half a wavelength
From (i),

$$\begin{aligned} \frac{V_g}{V_a} &= \frac{\lambda_g}{\lambda_a} \\ \therefore \frac{V_g}{V_a} &= \frac{d_b}{d_a} \quad \dots \quad (13) \end{aligned}$$

The two velocities can thus be compared as d_b , d_a are known; and if the velocity of sound, V_a , in air is known, the velocity in carbon dioxide can be calculated.

Vibrations in Plates. Chladni's Figures

We have already studied the different modes of vibration of the air in a pipe, the particles of a string, and the particles of a rod. About 1790 CHLADNI examined the vibrations of a glass *plate* by sprinkling sand on it. If the plate on a stand is gripped firmly at the corner N and bowed in the middle A of one side, the particles arrange themselves into a symmetrical pattern which shows the nodes of the stationary wave in the plate, Fig. 27.29 (i). By gripping the plate firmly at other points N, thus making a node at these points, and bowing at A, a series

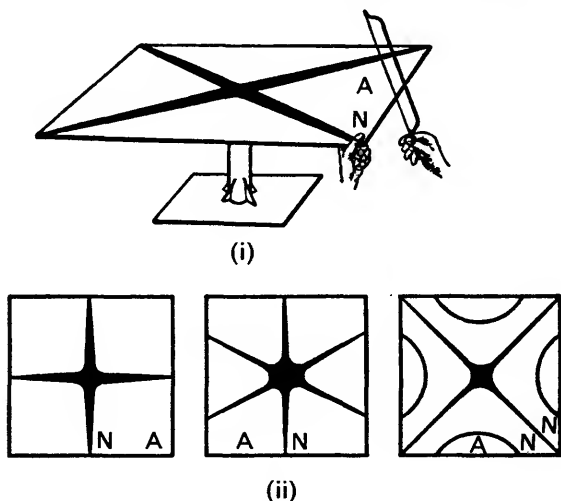


FIG. 27.29. Chladni's figures.

of different patterns can be obtained. These are known as *Chladni's figures*, Fig. 27.29 (ii). Each pattern corresponds to a particular mode of vibration. These modes are not harmonically related in frequency, unlike the case of the vibration of air in pipes and the vibration of strings.

EXAMPLES

1. Describe and explain the way in which a Kundt tube may be used to determine the ratio of the specific heats of a gas. A Kundt tube is excited by a brass rod 150 cm long and the distance between successive nodes in the tube is 13.6 cm; what is the ratio of the velocity of sound in brass to that in air? (*L.*)

First part. See text.

Second part. Since both ends of the rod are successive antinodes, the wavelength λ_1 in the rod = $2 \times 150 = 300$ cm. The wavelength λ_2 in the air = $2 \times 13.6 = 27.2$ cm.

The frequency f of the note in the rod and the air is the same.

$$\therefore f = \frac{V_1}{\lambda_1} = \frac{V_2}{\lambda_2}$$

where V_1, V_2 are the velocities of sound in the rod and in the air.

$$\therefore \frac{V_1}{V_2} = \frac{\lambda_1}{\lambda_2} = \frac{300}{27.2} = 11.0$$

2. Describe the dust tube experiment. How may it be used to compare the velocities of sound in different gases? The fundamental frequency of longitudinal vibration of a rod clamped at its centre is 1500 Hz. If the mass of the rod is 96.0 g, find the increase in its total length produced by a stretching force of 10 kgf (*L.*)

First part. The dust tube is Kundt's tube. See p. 666.

Second part. The wavelength of the wave in the rod = $2l$, where l metre is its length, since the ends are antinodes. The velocity V , of the wave is given by $V = f\lambda = 1500 \times 2l = 3000l$ (i)
Since the vibrations of the rod are longitudinal,

$$V = \sqrt{\frac{E}{\rho}}$$

E is Young's modulus in N m^{-2} and ρ is the density of the rod in kg m^{-3}

$$\therefore V = \sqrt{\frac{E}{0.096/v}} = \sqrt{\frac{Ev}{0.096}} \quad \text{. (ii)}$$

where v is the volume of the rod in metre^3 .

From (i) and (ii),

$$\sqrt{\frac{Ev}{0.096}} = 3000l$$

$$\therefore \frac{Ev}{l^2} = 0.096 \times 3000^2$$

$$\therefore \frac{EA}{l} = 0.096 \times 3000^2 \quad \text{. (iii)}$$

since $v = Al$, where A is the area of cross-section of the rod. Now if x is the increase in length produced by 10 kgf, it follows from the definition of E that

$$\text{force} = \frac{EAx}{l} = 10 \times 9.8 \text{ newtons}$$

$$\therefore \frac{EA}{l} = \frac{10 \times 9.8}{x}$$

$$\text{From (iii), } \therefore 0.096 \times 3,000^2 = \frac{10 \times 9.8}{x}$$

$$\therefore x = \frac{10 \times 9.8}{0.096 \times 3000^2} = 1.1 \times 10^{-4} \text{ metre.}$$

EXERCISES 27

Pipes

1. Write down in terms of wavelength, λ , the distance between (i) consecutive nodes, (ii) a node and an adjacent antinode, (iii) consecutive antinodes. Find the frequency of the fundamental of a closed pipe 15 cm long if the velocity of sound in air is 340 m s^{-1} .

2. Discuss what is meant by the statement that *sound is a wave motion*. Use the example of the passage of a sound wave through air to explain the terms wavelength (λ), frequency (f), and velocity (v) of a wave. Show that $v = f\lambda$.

Explain the increase in loudness (or 'resonance') which occurs when a sounding tuning-fork is held near the open end of an organ pipe when the length of the pipe has certain values, the other end of the pipe being closed. Find the shortest length of such a pipe which resonates with a 440 Hz tuning-fork, neglecting end corrections. (Velocity of sound in air = 350 m s^{-1} .) (O. & C.)

3. What are the chief characteristics of a progressive wave motion? Give your reasons for believing that sound is propagated through the atmosphere as a longitudinal wave motion, and find an expression relating the velocity, the frequency, and the wavelength.

Neglecting end effects, find the lengths of (a) a closed organ pipe, and (b) an open organ pipe, each of which emits a fundamental note of frequency 256 Hz. (Take the speed of sound in air to be 330 m s^{-1} .) (O.)

4. (a) Explain in terms of the properties of a gas, but without attempting mathematical treatment, how the vibration of a sound source, such as a loud-speaker diaphragm, can be transmitted through the air around it.

Explain, also, the reflection which occurs when the vibration reaches a fixed barrier, such as a wall.

(b) Plane, simple harmonic, progressive sound waves of wavelength 1.2 m and speed 348 m s^{-1} , are incident normally on a plane surface which is a perfect reflector of sound. What statements can be made about the amplitude of vibration and about air pressure changes at points distant (i) 30 cm, (ii) 60 cm, (iii) 90 cm, (iv) 10 cm from the reflector? Justify your answers. (O. & C.)

5. Describe the motion of the air in a tube closed at one end and vibrating in its fundamental mode. An observer (a) holds a vibrating tuning-fork over the open end of a tube which resounds to it, (b) blows lightly across the mouth of the tube. Describe and explain the difference in the quality of the notes that he hears.

A uniform tube, 60.0 cm long, stands vertically with its lower end dipping into water. When the length above water is 14.8 cm, and again when it is 48.0 cm, the tube resounds to a vibrating tuning fork of frequency 512 Hz. Find the lowest frequency to which the tube will resound when it is open at both ends. (L.)

6. Discuss the factors which determine the pitch of the note given by a 'closed' pipe. Explain why the fundamental frequency and the quality of the note from a 'closed' pipe differ from those of the note given under similar conditions by a pipe of the same length which is open at both ends. (N.)

7. What is meant by (a) a stationary wave motion and (b) a node?

Describe how the phenomenon of resonance may be demonstrated using a loudspeaker, a source of alternating voltage of variable frequency and a suitable tube open at one end and closed at the other. Explain how resonance occurs in the arrangement you describe, draw a diagram showing the position of the nodes in the tube in a typical case of resonance and state clearly the meaning of the diagram. How would you demonstrate the position of the nodes experimentally? (O. & C.)

8. Explain the meaning of (a) the *end correction* of a resonance tube, (b) *beats*. Establish a formula for the frequency of beats in terms of the superimposed frequencies.

A closed resonance tube with an end correction of 0.60 cm is made to sound its fundamental note on a day when the air temperature is 17°C . It is found to be in unison with a siren whose disc, which has 12 holes, is revolving at a rate of 43.0 rev s^{-1} . Calculate (i) the length of the tube, (ii) the frequency of the beats produced if the experiment is repeated on a day when the air temperature has fallen to 12°C , the rate of revolution of the siren's disc being unaltered. (The velocity of sound in air at 0°C may be taken as 331.5 m s^{-1} .) (L.)

9. Distinguish between the formation of an echo and the formation of a stationary sound wave by reflection, explaining the general circumstances in which each is produced.

Describe an experiment in which the velocity of sound in air may be determined by observations on stationary waves.

An organ pipe is sounded with a tuning-fork of frequency 256 Hz. When the air in the pipe is at a temperature of 15°C , 23 beats are heard in 10 seconds; when the tuning-fork is loaded with a small piece of wax, the beat frequency is found to decrease. What change of temperature of the air in the pipe is necessary to bring the pipe and the unloaded fork into unison? (C.)

10. Describe and give the theory of one experiment in each instance by which the velocity of sound may be determined, (a) in free air, (b) in the air in a resonance tube.

What effect, if any, do the following factors have on the velocity of sound in free air; frequency of the vibrations; temperature of the air; atmospheric pressure; humidity?

State the relationship between this velocity and temperature. (L.)

Strings. Rods

11. What is meant by a *wave motion*? Define the terms *wavelength* and *frequency* and derive the relationship between them.

Given that the velocity v of transverse waves along a stretched string is related to the tension F and the mass m per unit length by the equation

$$v = \sqrt{\frac{F}{m}},$$

derive an expression for the natural frequencies of a string of length l when fixed at both ends.

Explain how the vibration of a string in a musical instrument produces sound and how this sound reaches the ear. Discuss the factors which determine the quality of the sound heard by the listener. (*O. & C.*)

12. Distinguish between a *progressive* wave and a *stationary* wave. Explain in detail how you would use a sonometer to establish the relation between the fundamental frequency of a stretched wire and (a) its length, (b) its tension. You may assume a set of standard tuning-forks and a set of weights in steps of half a kilogram to be available.

A pianoforte wire having a diameter of 0.90 mm is replaced by another wire of the same material but with diameter 0.93 mm. If the tension of the wire is the same as before, what is the percentage change in the frequency of the fundamental note? What percentage change in the tension would be necessary to restore the original frequency? (*L.*)

13. What is meant by (a) a forced vibration, (b) resonance? Give an example of each from (i) mechanics, (ii) sound.

Using the same axes sketch graphs showing how the amplitude of a forced vibration depends upon the frequency of the applied force when the damping of the system is (a) light, (b) heavy. Point out any special features of the graphs.

A sonometer wire is stretched by hanging a metal cylinder of density 8000 kg m^{-3} at the end of the wire. A fundamental note of frequency 256 Hz is sounded when the wire is plucked.

Calculate the frequency of vibration of the same length of wire when a vessel of water is placed so that the cylinder is totally immersed. (*N.*)

14. Describe an experiment to determine the velocity of sound in a gas, e.g. nitrogen. How would you expect the velocity to be affected by (a) temperature, (b) pressure and (c) humidity? Give reasons for your answers.

What information about the nature of a gas can be obtained from a measurement of the velocity of sound in that gas, the pressure and density being known? (*L.*)

15. Describe experiments to illustrate the differences between (a) *transverse* waves, (b) *longitudinal* waves, (c) *progressive* waves and (d) *stationary* waves? To which classes belong (i) the vibrations of a violin string, (ii) the sound waves emitted by the violin into the surrounding air?

A wire whose mass per unit length is $10^{-3} \text{ kg m}^{-1}$ is stretched by a load of 4 kg over the two bridges of a sonometer 1 m apart. If it is struck at its middle point, what will be (a) the wavelength of its subsequent fundamental vibrations, (b) the fundamental frequency of the note emitted? If the wire were struck at a point near one bridge what further frequencies might be heard? (Do not derive standard formulae.) (Assume $g = 10 \text{ m s}^{-2}$.) (*O. & C.*)

16. A uniform wire vibrates transversely in its fundamental mode. On what factors, other than the length does the frequency of vibration depend, and what is the form of the dependence for each factor?

Describe the experiment you would perform to verify the form of dependence for *one* factor.

A wire of diameter 0.040 cm and made of steel of density 8000 kg m^{-3} is under constant tension of 8.0 kgf. A fixed length of 50 cm is set in transverse vibration. How would you cause the vibration of frequency about 840 Hz to predominate in intensity? (*N.*)

17. (a) The velocity of sound in air being known, describe how Young's modulus for brass may be found using Kundt's tube. (b) Discuss how the frequency of a note heard by an observer is affected by movement of (i) the source, (ii) the observer along the line joining source and observer. (*L.*)

18. Give an expression for the velocity of a transverse wave along a thin flexible string and show that it is dimensionally correct. Explain how reflexion may give rise to transverse *standing waves* on a stretched string and use the expression for the velocity to derive the frequency of the fundamental mode of vibration.

A steel wire of length 40.0 cm and diameter 0.0250 cm vibrates transversely in unison with a tube, open at each end and of effective length 60.0 cm, when each is sounding its fundamental note. The air temperature is 27°C . Find in kilograms force the tension in the wire. (Assume that the velocity of sound in air at 0°C is 331 m s^{-1} and the density of steel is 7800 kg m^{-3} .) (*L.*)

19. It may be shown theoretically that the frequency f of the fundamental note emitted as the result of the transverse vibration of a stretched wire is given by $f = kT^{1/2}l^{-1}$, where T is the tension in the wire and l its length, k being a constant for a given wire. Describe the experiments you would perform to check this relation, assuming that tuning-forks of known frequencies covering a range of one octave are available.

A brass wire is tuned so that its fundamental frequency is 100 Hz, and a horse-shoe magnet is placed so that the mid-point of the wire lies between its poles. On passing an alternating current through the wire it vibrates, the amplitude of vibration depending upon the frequency of the current. Explain this, and show for what frequencies the vibration will be particularly strong. (*C.*)

20. Explain why the velocity of sound in a gas depends upon the ratio of its principal specific heats.

Give a detailed account of a method of determining this ratio for carbon dioxide gas at atmospheric temperature, assuming that the ratio for air is known. Give the theory of the experiment. (*W.*)

OPTICS

chapter twenty-eight

Wave theory of light

Historical

IT has already been mentioned that light is a form of energy which stimulates our sense of vision. One of the early theories of light, about 400 B.C., suggested that particles were emitted from the eye when an object was seen. It was realised, however, that something is *entering* the eye when a sense of vision is caused, and about 1660 the great Newton proposed that particles, or corpuscles, were emitted from a luminous object. The *corpuscular theory of light* was adopted by many scientists of the day owing to the authority of Newton, but HUYGENS, an eminent Dutch scientist, proposed about 1680 that light energy travelled from one place to another by means of a wave-motion. If the *wave theory of light* was correct, light should bend round a corner, just as sound travels round a corner. The experimental evidence for the wave theory in Huygens' time was very small, and the theory was dropped for more than a century. In 1801, however, THOMAS YOUNG obtained evidence that light could produce wave effects (p. 688), and he was among the first to see clearly the close analogy between sound and light waves. As the principles of the subject became understood other experiments were carried out which showed that light could spread round corners, and Huygens' wave-theory of light was revived. Newton's corpuscular theory was rejected since it was incompatible with experimental observations (see p. 679), and the wave theory of light has played, and is still playing, an important part in the development of the subject.

In 1905 the great mathematical physicist EINSTEIN suggested that the energy in light could be carried from place to place by particles whose energy depended on the wavelength of the light. This was a return to a corpuscular theory, though it was completely different from that of Newton, as we see later. Experiments carried out at his suggestion showed that the theory was true, and the particles of light energy are known as "photons" (p. 1080). It is now considered that *either* the wave theory *or* the particle theory of light can be used in a problem on light, depending on the circumstances of the problem. In this book we shall first consider Huygens' wave theory, which was the foundation of many notable advances in the subject.

Wavefront. Rays

We have already considered the topic of waves in the Sound section

(p. 583). As we shall presently see, close analogies exist between light and sound waves.

Consider a point source of light, S , in air, and suppose that a disturbance, or wave, originates at S as a result of vibrations occurring inside the atoms of the source, and travels outwards. After a time t the wave has travelled a distance ct , where c is the velocity of light in air, and the light energy has thus reached the surface of a sphere of centre S and radius ct , Fig. 28.1. The surface of the sphere is called the *wavefront* of the light at this instant, and every point on it is vibrating "in step" or *in phase* with every other point. As time goes on the wave travels further and new wavefronts are obtained which are the surfaces of spheres of centre S .

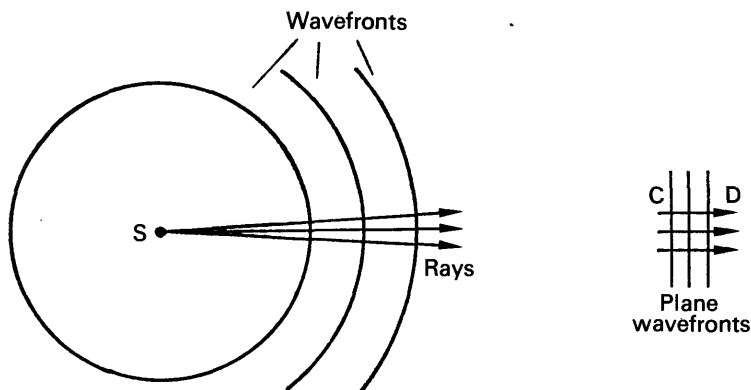


FIG. 28.1. Wavefronts and rays.

At points a long way from S , such as C or D , the wavefronts are portions of a sphere of very large radius, and the wavefronts are then substantially *plane*. Light from the sun reaches the earth in plane wavefronts because the sun is so far away; plane wavefronts also emerge from a convex lens when a point source of light is placed at its focus.

The significance of the wavefront, then, is that it shows how the light energy travels from one place in a medium to another. A *ray* is the name given to the direction along which the energy travels, and consequently a ray of light passing through a point is perpendicular to the wavefront at that point. The rays diverge near S , but they are approximately parallel a long way from S , as plane wavefronts are then obtained, Fig. 28.1.

Huygens' Construction for the New Wavefront

Suppose that the wavefront from a centre of disturbance S has reached the surface AB in a medium at some instant, Fig. 28.2. To obtain the position of the new wavefront after a further time t , Huygens postulated that *every point, $A, \dots, C, \dots, E, \dots, B$, on AB becomes a new or "secondary" centre of disturbance*. The wavelet from A then reaches the surface M of a sphere of radius vt and centre A , where v is the velocity

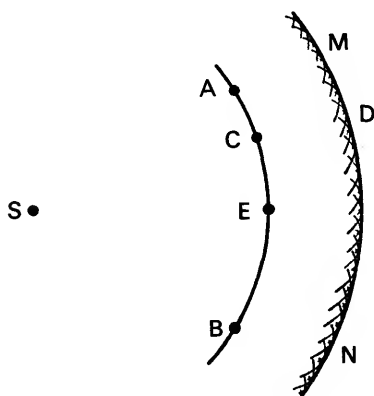


FIG. 28.2. Huygens' construction.

of light in the medium; the wavelet from C reaches the surface D of a sphere of radius vt and centre C; and so on for every point on AB. According to Huygens, *the new wavefront is the surface MN which touches all the wavelets from the secondary sources*; and in the case considered, it is the surface of a sphere of centre S.

In this simple example of obtaining the new wavefront, the light travels in the same medium. Huygens' construction, however, is especially valuable for deducing the new wavefront when the light travels from one medium to another, as we shall soon show.

Reflection at Plane Surface

Suppose that a beam of parallel rays between HA and LC is incident on a plane mirror, and imagine a plane wavefront AB which is normal to the rays, reaching the mirror surface, Fig. 28.3. At this instant the point A acts as a centre of disturbance. Suppose we require the new wavefront at a time corresponding to the instant when the disturbance at B reaches C. The wavelet from A reaches the surface of a sphere of radius AD at this instant; and as other points between AC on the mirror, such as P, are reached by the disturbances originating on AB, wavelets

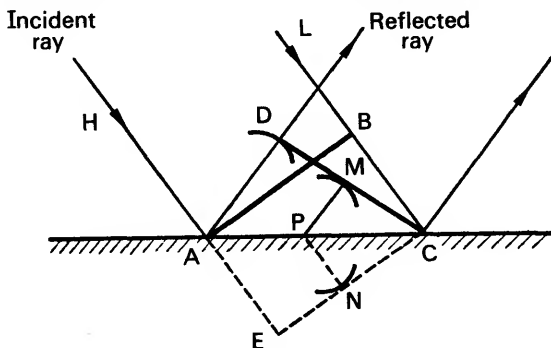


FIG. 28.3. Reflection at plane surface.

of smaller radius than AD are obtained at the instant we are considering. The new wavefront is the surface CMD which touches all the wavelets.

In the absence of the mirror, the plane wavefront AB would reach the position EC in the time considered. Thus $AD = AE = BC$, and $PN = PM$, where PN is perpendicular to EC. The triangles PMC, PNC are hence congruent, as PC is common, angles PMC, PNC are each 90° , and $PN = PM$. Thus angle PCM = angle PCN. But triangles ACD, AEC are also congruent. Consequently angle ACD = angle ACE = angle PCN = angle PCM. Since EC is a plane. Hence CMD is a *plane* surface.

Law of reflection. We can now deduce the law of reflection concerning the angles of incidence and reflection. From the above, it can be seen that the triangles ABC, AEC are congruent, and that triangles ADC, AEC are congruent. The triangles ABC, ADC are hence congruent, and therefore angle BAC = angle DCA. Now these are the angles made by the wavefront AB, CD respectively with the mirror surface AC. Since the incident and reflected rays, for example HA, AD, are normal to the wavefronts, these rays also make equal angles with AC. It now follows that the angles of incidence and reflection are equal.

Point Object

Consider now a point object O in front of a plane mirror M, Fig. 28.4. A spherical wave spreads out from O, and at some time the wavefront reaches ABC. In the absence of the mirror the wavefront would reach a position DEF in a time t thereafter, but every point between D and F on the mirror acts as a secondary centre of disturbance and wavelets are reflected back into the air. At the end of the time t , a surface DGC is drawn to touch all the wavelets, as shown. DGC is part of a spherical surface which advances into the air, and it appears to have come from a point I as a centre below the mirror, which is therefore a virtual image.

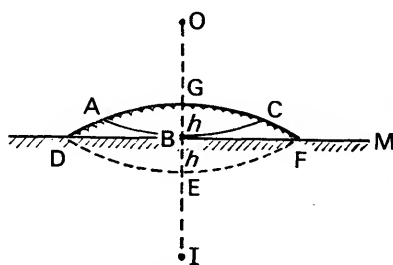


FIG. 28.4. Point object.

The sphere of which DGC is part has a chord DF. Suppose the distance from B, the midpoint of the chord, to G is h . The sphere of which DEF is part has the same chord DF, and the distance from B to E is also h . It follows, from the theorem of product of intersection of chords of a circle, that $DB \cdot BF = h(2r - h) = h(2R - h)$, where r is the radius OE and R is the radius IG. Thus $R = r$, or $IG = OE$, and hence $IB = OB$. The image and object are thus equidistant from the mirror.

Refraction at Plane Surface

Consider a beam of parallel rays between LO and MD incident on the plane surface of a water medium from air in the direction shown, and suppose that a plane wavefront has reached the position OA at a certain

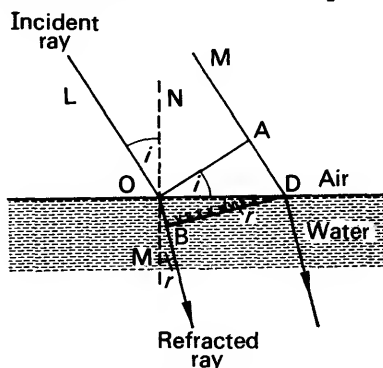


FIG. 28.5. Refraction at plane surface.

instant, Fig. 28.5. Each point between O, D becomes a new centre of disturbance as the wavefront advances to the surface of the water, and the wavefront changes in direction when the disturbance enters the liquid.

Suppose that t is the time taken by the light to travel from A to D. The disturbance from O travels a distance OB, or vt , in water in a time t , where v is the velocity of light in water. At the end of the time t , the wavefronts in the water from the other secondary centres between O, D reach the surfaces of spheres to each of which DB is a tangent. Thus DB is the new wavefront in the water, and the ray OB which is normal to the wavefront is consequently the refracted ray.

Since c is the velocity of light in air, $AD = ct$. Now

$$\begin{aligned} \frac{\sin i}{\sin r} &= \frac{\sin \angle LON}{\sin \angle BOM} = \frac{\sin \angle AOB}{\sin \angle ODB} \\ \therefore \frac{\sin i}{\sin r} &= \frac{AD/OD}{OB/OD} = \frac{AD}{OB} = \frac{ct}{vt} = \frac{c}{v} \quad \dots \quad (i) \end{aligned}$$

But c, v are constants for the given media.

$$\therefore \frac{\sin i}{\sin r} \text{ is a constant,}$$

which is Snell's law of refraction (p. 420).

It can now be seen from (i) that the refractive index, n , of a medium is given by $n = \frac{c}{v}$, where c is the velocity of light *in vacuo* and v is the velocity of light in the medium.

Newton's Corpuscular Theory of Light

Prior to the wave theory of light, Newton had proposed a corpuscular or particle theory of light. According to Newton, particles are emitted

by a source of light, and they travel in a straight line until the boundary of a new medium is encountered.

In the case of *reflection at a plane surface*, Newton stated that at some very small distance from the surface M, represented by AB, the particles were acted upon by a repulsive force, which gradually diminished the component of the velocity v in the direction of the normal and then reversed it, Fig. 28.6. The *horizontal* component of the velocity remained unaltered, and hence the velocity of the particles of light as they moved away from M is again v . Since the horizontal components of the incident and reflected velocities are the same, it follows that

$$v \sin i = v \sin i' \quad \dots \quad (i)$$

where i' is the angle of reflection.

$$\therefore \sin i = \sin i', \text{ or } i = i'$$

Thus the corpuscular theory explains the law of reflection at a plane surface.

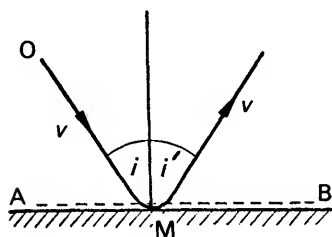


FIG. 28.6. Newton's corpuscular theory of reflection.

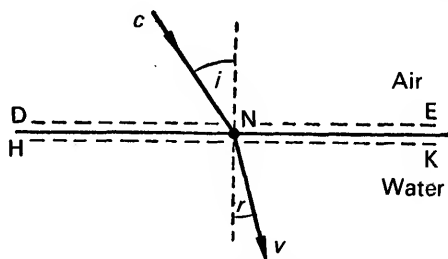


FIG. 28.7. Newton's corpuscular theory of refraction.

To explain *refraction at a plane surface* when light travels from air to a denser medium such as water, Newton stated that a force of attraction acted on the particles as they approached beyond a line DE very close to the boundary N, Fig. 28.7. The vertical component of the velocity of the particles was thus increased on entering the water, the horizontal component of the velocity remaining unaltered, and beyond a line HK close to the boundary the vertical component remained constant at its increased value. The resultant velocity, v , of the particles in the water is thus *greater* than its velocity, c , in air.

Suppose i , r are the angles of incidence and refraction respectively. Then, as the horizontal components of the velocity is unaltered.

$$c \sin i = v \sin r$$

$$\therefore \frac{\sin i}{\sin r} = \frac{v}{c}$$

$$\therefore n = \frac{v}{c} = \text{the refractive index}$$

Since n is greater than 1, the velocity of light (v) in water is greater than the velocity (c) in air, as was stated above. This is according to Newton's corpuscular theory. On the wave theory, however, $n = \frac{c}{v}$ (see p. 679);

and hence the velocity of light (v) in water is *less* than the velocity (c) in air according to the wave theory. The corpuscular theory and wave theory are thus in conflict, and Foucault's experimental results showed that the corpuscular theory, as enunciated by Newton, could not be true (p. 559).

Dispersion

The dispersion of colours produced by a medium such as glass is due to the difference in speeds of the various colours in the medium. Thus suppose a plane wavefront AC of white light is incident in air on a plane glass surface, Fig. 28.8. In the time the light takes to travel in air from C to D , the red light from the centre of disturbance A reaches a position shown by the wavelet at R . The blue light from A reaches another position shown by the wavelet at B , since the speed of blue light in glass is less than that of red light, so that AB is less than AR . On drawing the new wavefronts DB , DR , it can be seen that the blue wavefront BD is refracted more in the glass than the red wavefront DR . The refracted blue ray is AB and the refracted red ray is AR , and hence dispersion occurs.

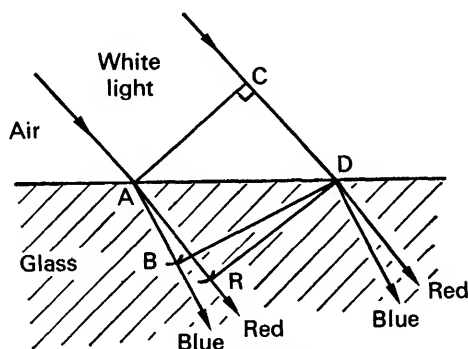


FIG. 28.8. Dispersion.

Refraction Through Prism at Minimum Deviation

Consider a wavefront HB incident on the face HB of a prism of angle A , Fig. 28.9. If the wavefront emerges along EC , then the light travels

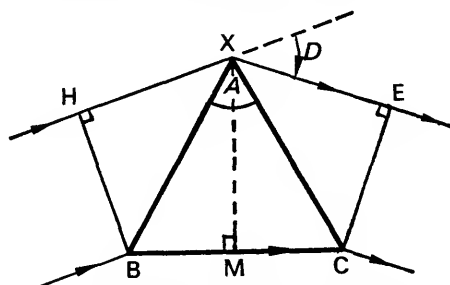


FIG. 28.9. Refraction at minimum deviation.

a distance HXE in air in the same time as the light travels a distance BC in glass.

$$\therefore \frac{HX + XE}{c} = \frac{BC}{v},$$

where c is the velocity of light in air and v is the velocity in glass

$$\therefore HX + XE = \frac{c}{v} BC = n BC \quad \text{. (i)}$$

At minimum deviation, the wavefront passes symmetrically through the prism (p. 443).

$$\therefore HX = XE$$

From (i),

$$\therefore 2 HX = n BC$$

$$\therefore n = \frac{2 HX}{BC} \quad \text{. (ii)}$$

$$\begin{aligned} \text{But } HX &= XB \cos BXH = XB \cos \left[\frac{180^\circ - (A + D)}{2} \right] \\ &= XB \sin \left(\frac{A + D}{2} \right), \end{aligned}$$

$$\text{and} \quad BC = 2 BM = 2 XB \sin \frac{A}{2}$$

$$\text{From (ii),} \quad \therefore n = \frac{\sin \left(\frac{A + D}{2} \right)}{\sin \frac{A}{2}}$$

Focal Length of Lens

The focal length of a lens (or curved spherical mirror) can also be found by wave theory. Suppose a plane wavefront AHB, parallel to

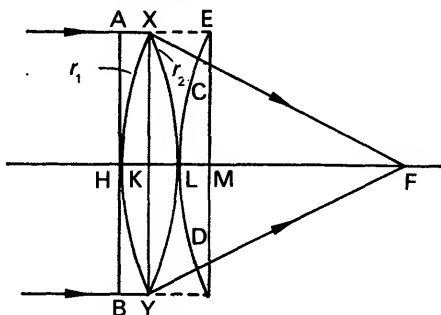


FIG. 28.10. Focal length of lens.

the principal axis, is incident on a converging lens, Fig. 28.10. After refraction the wavefront emerges in air as a converging spherical wavefront CLD of centre F, the principal focus, since the incident rays are parallel.

The time taken by the light to travel a distance $AX + XE$ in air is equal to the time taken to travel a distance HKL in glass. Thus if c is the velocity in air and v is the velocity in glass,

$$\therefore \frac{AX + XE}{c} = \frac{HKL}{v}$$

$$\therefore AX + XE = \frac{c}{v} \cdot HKL = n \cdot HKL \quad (i)$$

From the geometry of a circle, $AX = HK = \frac{h^2}{2r_1}$,

where XK is h and r_1 is the radius of curvature of the lens, assumed thin (see also p. 695).

Also,
$$XE = KM = KL + LM = \frac{h^2}{2r_2} + \frac{h^2}{2f},$$

where r_2 is the radius of curvature of the surface XY and $FL = f$. Substituting in (i),

$$\therefore \frac{h^2}{2r_1} + \frac{h^2}{2r_2} + \frac{h^2}{2f} = n \left(\frac{h^2}{2r_1} + \frac{h^2}{2r_2} \right)$$

Simplifying,
$$\therefore \frac{1}{f} = (n - 1) \left(\frac{1}{r_1} + \frac{1}{r_2} \right).$$

Power of a Lens

We have now to consider the effect of lenses on the *curvature* of wavefronts. A plane wavefront has obviously zero curvature and a spherical wavefront has a small curvature if the radius of the sphere is large. The "*curvature*" of a spherical wavefront is defined as $1/r$, where r is the radius of the surface which constitutes the wavefront, and hence the curvature is zero when r is infinitely large, as in the case of a plane wavefront.

When a plane wavefront is incident on a converging lens L , a spherical wavefront, S , of radius f emerges from L , where f is the focal length of the lens, Fig. 28.11 (i). Parallel rays, which are normal to the plane

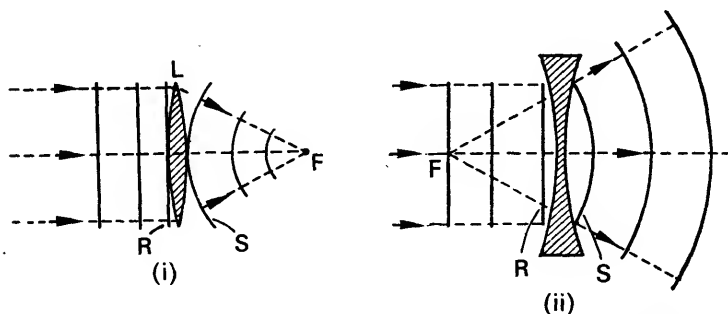


FIG. 28.11. (i). Converging lens.

(ii). Diverging lens.

wavefront, are thus refracted towards F , the focus of the lens. Now the curvature of a plane wavefront is zero, and the curvature of the spherical wavefront S is $1/f$. Thus the convex lens impresses a curvature

of $1/f$ on a wavefront incident on it, and $1/f$ is accordingly defined as the *converging power* of the lens.

$$\therefore \text{Power } P, = \frac{1}{f} \quad (91)$$

Fig. 28.11 (ii) illustrates the effect of a *diverging* lens on a plane wavefront R. The front S emerging from the lens has a curvature opposite to S in Fig. 28.11 (i), and it appears to be diverging from a point F behind the concave lens, which is its focus. The curvature of the emerging wavefront is thus $1/f$, where f is the focal length of the lens, and the powers of the convex and concave lens are opposite in sign.

The power of a converging lens is positive, since its focal length is positive, while the power of a diverging lens is negative. Opticians use a unit of power called the *diopetre*, D , which is defined as the power of a lens of 100 cm focal length. A lens of focal length f cm has thus a power P given by

$$P = \frac{1/f}{1/100} \text{ dioptries}$$

$$\text{or } P = \frac{100}{f} \text{ dioptries} \quad (92)$$

A lens of $+8$ dioptries, or $+8D$, is therefore a converging lens of focal length 12.5 cm, and a lens of $-4D$ is a diverging lens of 25 cm focal length.

The Lens Equation

Suppose that an object O is placed a distance u from a converging lens, Fig. 28.12. The spherical wavefront A from O which reaches the lens has a radius of curvature u , and hence a curvature $1/u$. Since the converging lens adds a curvature of $1/f$ to the wavefront (proved page 683), the spherical wavefront B emerging from

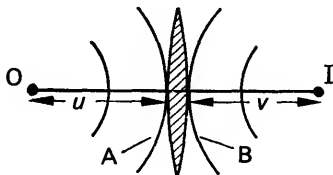


FIG. 28.12. Effect of lens on wavefront.

the lens into the air has a curvature $\left(\frac{1}{u} + \frac{1}{f}\right)$. But the curvature is also given by $\frac{1}{v}$, where v is the image distance IB from the lens.

$$\therefore \frac{1}{v} = \frac{1}{u} + \frac{1}{f}$$

It can be seen that the curvature of A is of an opposite sign to that of B; and taking this into account, the lens equation $\frac{1}{v} + \frac{1}{u} = \frac{1}{f}$ is obtained.

A similar method can be used for a diverging lens, which is left as an exercise for the student.

EXERCISES 28

1. A parallel beam of monochromatic radiation travelling through glass is incident on the plane boundary between the glass and air. Using Huygens' principle draw diagrams (one in each case) showing successive positions of the wave fronts when the angle of incidence is (a) 0° , (b) 30° , (c) 60° . Indicate clearly and explain the constructions used. (The refractive index of glass for the radiation used is 1.5.) (N.)

2. State Snell's law of refraction. How is the law explained in terms of the wave theory of light?

An equiangular glass prism is placed in a broad beam of parallel monochromatic light as shown, the face AB of the prism being perpendicular to the direction of the incident light, Fig. 28A. By sketching typical rays, show that most of the light which is refracted by the prism emerges as two beams of parallel light deviated respectively through $\pm \theta$, and calculate the value of θ if the refractive index of the glass is 1.5. Why does not all the light falling on the prism emerge in this way?

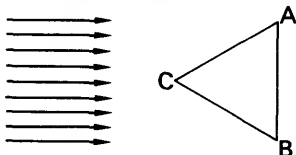


FIG. 28A.

The prism is turned round so that the light is incident normally on the face AB. Describe as fully as you can what happens to most of the light with the prism in this position. (O. & C.)

3. A plane wave-front of monochromatic light is incident normally on one face of a glass prism, of refracting angle 30° , and is transmitted. Using Huygens' construction trace the course of the wave-front. Explain your diagram and find the angle through which the wave-front is deviated. (Refractive index of glass = 1.5.) (N.)

4. State *Snell's law of refraction* and define *refractive index*.

Show how refraction of light at a plane interface can be explained on the basis of the wave theory of light.

Light travelling through a pool of water in a parallel beam is incident on the horizontal surface. Its speed in water is $2.2 \times 10^8 \text{ m s}^{-1}$. Calculate the maximum angle which the beam can make with the vertical if light is to escape into the air where its speed is $3.0 \times 10^8 \text{ m s}^{-1}$.

At this angle in water, how will the path of the beam be affected if a thick layer of oil, of refractive index 1.5, is floated on to the surface of the water? (O. & C.)

5. How did Huygens explain the reflection of light on the wave theory? Using Huygens' conceptions, show that a series of light waves diverging from a point source will, after reflection at a plane mirror, appear to be diverging from a second point, and calculate its position. (C.)

6. Explain how the *corpuscular theory* of Newton accounted for the laws of reflection and refraction. What experimental evidence showed that the theory was incorrect?

7. What is Huygens' principle?

Draw and explain diagrams which show the positions of a light wave-front at successive equal time intervals when (a) parallel light is reflected from a plane mirror, the angle of incidence being about 60° , (b) monochromatic light originating from a small source in water is transmitted through the surface of the water into the air.

Describe an experiment, and add the necessary theoretical explanation, to show that in air the wavelength of blue light is less than that of red light. (N.)

8. Using Huygens' concept of secondary wavelets show that a plane wave of monochromatic light incident obliquely on a plane surface separating air from glass may be refracted and proceed as a plane wave. Establish the physical significance of the refractive index of the glass.

In what circumstances does dispersion of light occur? How is it accounted for by the wave theory?

If the wavelength of yellow light in air is 6.0×10^{-7} m, what is its wavelength in glass of refractive index 1.5? (N.)

9. Describe fully a method for measuring the velocity of light in air. Explain, on the basis of the wave theory, the relation between the refractive index of a medium relative to air and the velocity of light. (L.)

chapter twenty-nine

Interference, diffraction, polarisation of light

INTERFERENCE OF LIGHT

THE beautiful colours seen in thin films of oil in the road, or in soap bubbles, are due to a phenomenon in light called *interference*. Newton discovered that circular coloured rings were obtained when white light illuminated a convex lens of large radius of curvature placed on a sheet of plane glass (p. 693), which is another example of interference. As we saw in Sound, interference can be used to measure the wavelength of sound waves (p. 617). By a similar method the phenomenon can be used to measure the wavelengths of different colours of light. Interference of light has also many applications in industry.

The essential conditions, and features, of interference phenomena have already been discussed in connection with sound waves. As there is an exact analogy between the interference of sound and light waves we can do no better than recapitulate here the results already obtained on pp. 616–617:

1. Permanent interference between two sources of light can only take place if they are *coherent* sources, i.e., they must have the same frequency and be always in phase with each other or have a constant phase difference. (This implies that the two sources of light must have the same colour.)

2. If the coherent monochromatic light sources are P, Q, a bright light is observed at B if the path-difference, $QB - PB$, is a whole number of wavelengths, Fig. 29.1. (This corresponds to the case of a loud sound

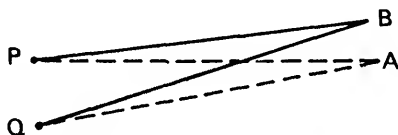


FIG. 29.1. Interference of light.

heard at B if P, Q were two coherent sources of sound.) A bright light is observed at A if $PA = QA$, in which case the path-difference is zero.

3. If the path-difference is an odd number of half wavelengths, darkness is observed at the point under consideration. (This corresponds to silence at the point in the case of two coherent sound sources.)

Young's Experiment

From the preceding, it can be understood that two conditions are essential to obtain an interference phenomenon. (i) Two coherent sources of light must be produced, (ii) the coherent sources must be very close to each other as the wavelength of light is very small, otherwise the bright and dark pattern in front of the sources tend to be too fine to see and no interference pattern is obtained.

One of the first demonstrations of the interference of light waves was given by YOUNG in 1801. He placed a source, S, of monochromatic light in front of a narrow slit C, and arranged two very narrow slits A, B, close to each other, in front of C. Much to his delight, Young observed bright and dark bands on either side of O on a screen T, where O is on the perpendicular bisector of AB, Fig. 29.2.

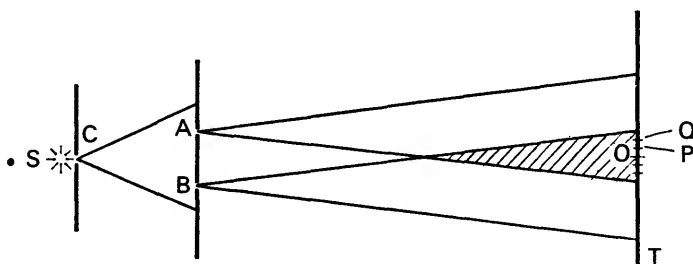


FIG. 29.2. Young's experiment.

Young's observations can be explained by considering the light from S illuminating the two slits A, B. Since the light diverging from A has exactly the same frequency as, and is always in phase with, the light diverging from B, A and B act as two close coherent sources. Interference thus takes place in the shaded region, where the light beams overlap, Fig. 29.2. As $AO = OB$, a bright band is obtained at O. At a point P close to O, such that $BP - AP = \lambda/2$, where λ is the wavelength of the light from S, a dark band is obtained. At a point Q such that $BQ - AQ = \lambda$, a bright band is obtained; and so on for either side of O. Young demonstrated that the bands were due to interference by covering A or B, when the bands disappeared.

Separation of Bands

Suppose P is the position of the m th bright band, so that $BP - AP = m\lambda$, Fig. 29.3. Let $OP = x_m$ = distance from P to O, the centre of the band system, where MO is the perpendicular bisector of AB. If a length PN equal to PA is described on PB, then $BN = BP - AP = m\lambda$. Now in practice AB is very small, and PM is very much larger than AB. Thus AN meets PM practically at right angles. It then follows that

angle PMO = angle BAN = θ say.

From triangle BAN, $\sin \theta = \frac{BN}{AB} = \frac{m\lambda}{a}$,

where $a = AB =$ the distance between the slits. From triangle PMO,

$$\tan \theta = \frac{PO}{MO} = \frac{x_m}{D},$$

where $D = MO =$ the distance from the screen to the slits. Since θ is very small, $\tan \theta = \sin \theta$.

$$\therefore \frac{x_m}{D} = \frac{m\lambda}{a}$$

$$\therefore x_m = \frac{mD\lambda}{a}$$

If Q is the neighbouring or $(m - 1)$ th bright band, it follows that

$$OQ = x_{m-1} = \frac{(m-1)D\lambda}{a}$$

$$\therefore \text{separation } y \text{ between successive bands} = x_m - x_{m-1} = \frac{\lambda D}{a} \quad (\text{i})$$

$$\therefore \lambda = \frac{ay}{b} \quad \dots \dots \dots \text{(ii)}$$

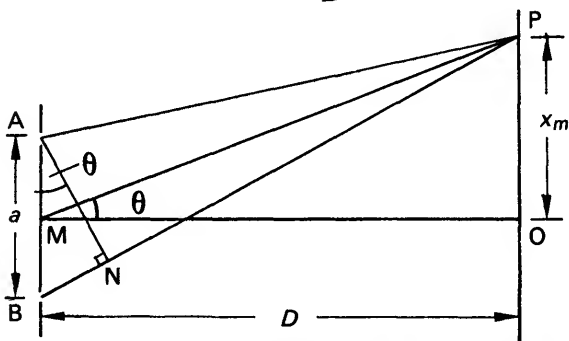


FIG. 29.3. Theory of Young's experiment (exaggerated).

Measurement of Wavelength by Young's Interference Bands

A laboratory experiment to measure wavelength by Young's interference bands is shown in Fig. 29.4. Light from a small filament lamp is focused by a lens on to a narrow slit S, such as that in the collimator of a spectrometer. Two narrow slits A, B, about a millimetre apart, are placed a short distance in front of S, and the light coming from A, B is viewed in a low-powered microscope or eyepiece M about two metres away. Some coloured interference bands are then observed by M. A red and then a blue filter, F, placed in front of the slits, produces red and then blue bands. Observation shows that the separation of the red bands is more than that of the blue bands. Now $\lambda = ay/D$, from (ii), where y is the separation of the bands. It follows that the wavelength of red light is *longer* than that of blue light.

An approximate value of the wavelength of red or blue light can be found by placing a Perspex rule R in front of the eyepiece and moving it until the graduations are clearly seen, Fig. 29.4. The average distance, y , between the bands is then measured on R. The distance a between

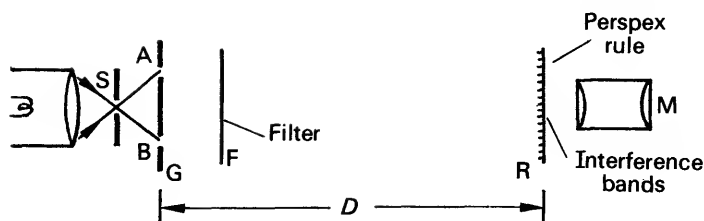


FIG. 29.4. Laboratory experiment on Young's interference bands.

the slits can be found by magnifying the distance by a convex lens, or by using a travelling microscope. The distance D from the slits to the Perspex rule, where the bands are formed, is measured with a metre rule. The wavelength λ can then be calculated from $\lambda = ay/D$, and is of the order 6×10^{-5} cm. Further details of the experiment can be obtained from *Advanced Level Practical Physics* by Nelkon and Ogborn (Heinemann).

The wavelengths of the extreme colours of the visible spectrum vary with the observer. This may be 4×10^{-5} cm for violet and 7×10^{-5} cm for red; an "average" value for visible light is 5.5×10^{-5} cm, which is a wavelength in the green.

Appearance of Young's Interference Bands

The experiment just outlined can also be used to demonstrate the following points:—

1. If the source slit S is moved nearer the double slits the separation of the bands is unaffected but their intensity increases. This can be seen from the formula y (separation) $= \lambda D/a$, since D and a are constant.

2. If the distance apart a of the slits is diminished, keeping S fixed, the separation of the bands increases. This follows from $y = \lambda D/a$.

3. If the source slit S is widened the bands gradually disappear. The slit S is then equivalent to a large number of narrow slits, each producing its own band system at different places. The bright and dark bands of different systems therefore overlap, giving rise to uniform illumination. It can be shown that, to produce interference bands which are recognisable, the slit width of S must be less than $\lambda D'/a$, where D' is the distance of S from the two slits A, B .

4. If one of the slits, A or B , is covered up, the bands disappear.

5. If white light is used the central band is white, and the bands either side are coloured. Blue is the colour nearer to the central band and red is farther away. The path difference to a point O on the perpendicular bisector of the two slits A, B is zero for all colours, and consequently each colour produces a bright band here. As they overlap, a white band is formed. Farther away from O , in a direction parallel to the slits, the shortest visible wavelengths, blue, produce a bright band first.

Fresnel's Biprism Experiment

Fresnel used a biprism R which had a very large angle of nearly 180° , and placed a narrow slit S , illuminated by monochromatic light, in

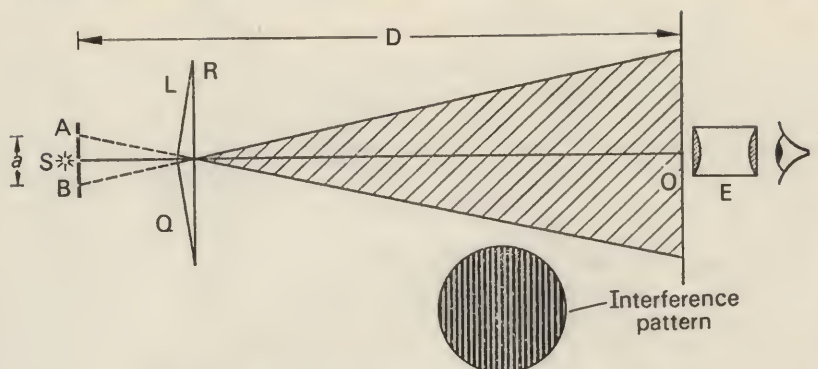


FIG. 29.5. Fresnel's biprism experiment (not to scale).

front of it so that the refracting edge was parallel to the slit, Fig. 29.5. The light emerging after refraction from the two halves, L, Q, of the prism can be considered to come from two sources, A, B, which are the virtual images of the slit S in L, Q respectively. Thus A, B are coherent sources; further, as R has a very large obtuse angle, A and B are close together. Thus an interference pattern is observed in the region of O where the emergent light from the two sources overlap, as shown by the shaded portion of Fig. 29.5, and bright and dark bands can be seen through an eyepiece E at O directed towards R, Fig. 29.6. By using cross-wires, and moving the eyepiece by a screw arrangement, the distance y between successive bright bands can be measured. Now it was shown on p. 689 that $\lambda = ay/D$, where a is the distance between A, B and D is the distance of the source slit from the eyepiece. The distance D is measured with a metre rule. The distance a can be found by moving a convex lens between the fixed biprism and eyepiece until a magnified image of the two slits A, B is seen clearly, and the magnified distance b between them is measured. The magnification m is (image distance \div object distance) for the lens, and a can be calculated from $a = b/m$. Knowing a , y , D , the wavelength λ can be determined.



FIG. 29.6 Fresnel's biprism interference bands (magnified).

If A is the large angle, nearly 180° , of the biprism, each of the small base angles is $(180^\circ - A)/2$, or $90^\circ - A/2$. The small deviation d in radians of light from the slit S is $(n - 1)\theta$, where θ is the magnitude of the base angle in radians (p. 457), and hence the distance A, B between the virtual images of the slit $= 2td = 2t(n - 1)\theta$, where t is the distance from S to the biprism.

Interference in Thin Wedge Films

A very thin wedge of an air film can be formed by placing a thin piece of foil or paper between two microscope slides at one end Y, with the slides in contact at the other end X, Fig. 29.7. The wedge has then a very small angle θ , as shown. When the air-film is illuminated by monochromatic light from an extended source S, straight bright and dark bands are observed which are parallel to the line of intersection X of the two slides.

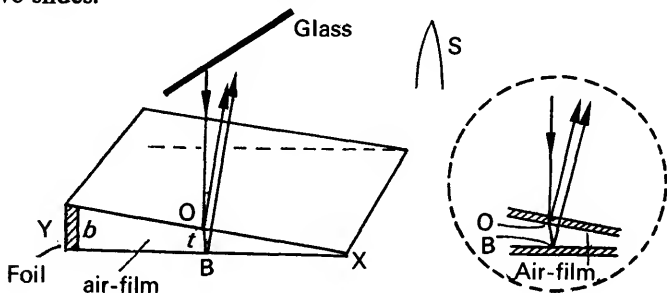


FIG. 29.7. Thin wedge film.

The light reflected down towards the wedge is partially reflected upwards from the lower surface O of the top slide. The remainder of the light passes through the slide and some is reflected upward from the top surface B of the lower slide. The two trains of waves are coherent, since both have originated from the same centre of disturbance at O, and they produce an interference phenomenon if brought together by the eye or in an eyepiece. Their path difference is $2t$, where t is the small thickness of the air-film at O. At X, where the path difference is apparently zero, we would expect a bright band. But a *dark* band is observed at X. This is due to a phase change of 180° , equivalent to an extra path difference of $\lambda/2$, which occurs when a wave is reflected at a denser medium. See pp. 641, 694. The optical path difference between the two coherent beams is thus actually $2t + \lambda/2$, and hence, if the beams are brought together to interfere, a bright band is obtained when $2t + \lambda/2 = m\lambda$, or $2t = (m - \frac{1}{2})\lambda$. A dark band is obtained at a thickness t given by $2t = m\lambda$.

The bands are located at the air-wedge film, and the eye or microscope must be focused here to see them. The appearance of a band is the contour of all points in the air-wedge film where the optical path difference is the same. If the wedge surfaces make perfect optical contact at one edge, the bands are straight lines parallel to the line of intersection of the surfaces. If the glass surfaces are uneven, and the contact at one edge is not regular, the bands are not perfectly straight. A particular band still shows the locus of all points in the air-wedge which have the same optical path difference in the air-film.

In *transmitted light*, the appearance of the bands are complementary to those seen by reflected light, from the law of conservation of energy. The bright bands thus correspond in position to the dark bands seen by reflected light, and the band where the surfaces touch is now bright instead of dark.

Thickness of Thin Foil. Expansion of Crystal

If there is a bright band at Y at the edge of the foil, Fig. 29.7, the thickness b of the foil is given by $2b = (m + \frac{1}{2})\lambda$, where m is the number of bright bands between X and Y. If there is a dark band at Y, then $2b = m\lambda$. Thus by counting m , the thickness b can be found. The small angle θ of the wedge is given by b/a , where a is the distance XY, and by measuring a with a travelling microscope focused on the air-film, θ can be found. If a liquid wedge is formed between the plates, the optical path difference becomes $2nt$, where the air thickness is t , n being the refractive index of the liquid. An optical path difference of λ now occurs for a change in t which is n times less than in the case of the air-wedge. The spacing of the bright and dark bands is thus n times closer than for air, and measurement of the relative spacing enables n to be found.

The coefficient of expansion of a crystal can be found by forming an air-wedge of small angle between a fixed horizontal glass plate and the upper surface of the crystal, and illuminating the wedge by monochromatic light. When the crystal is heated a number of bright bands, m say, cross the field of view in a microscope focused on the air-wedge. The increase in length of the crystal in an upward direction is $m\lambda/2$, since a change of λ represents a change in the thickness of the film is $\lambda/2$, and the coefficient of expansion can then be calculated.

Newton's Rings

Newton discovered an example of interference which is known as "Newton's rings". In this case a lens L is placed on a sheet of plane glass, G , having a lower surface of very large radius of curvature, Fig. 29.8. By means of a sheet of glass G monochromatic light from a sodium

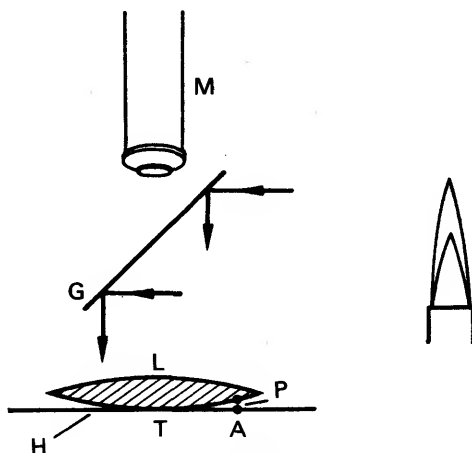


FIG. 29.8. Newton's rings.

flame, for example, is reflected downwards towards L ; and when the light reflected upwards is observed through a microscope M focused on

H, a series of bright and dark rings is seen. The circles have increasing radius, and are concentric with the point of contact T of L with H.

Consider the air-film PA between A on the plate and P on the lower lens surface. Some of the incident light is reflected from P to the microscope, while the remainder of the light passes straight through to A, where it is also reflected to the microscope and brought to the same focus. The two rays of light have thus a net path difference of $2t$, where $t = PA$. The same path difference is obtained at all points round T which are distant TA from T; and hence if $2t = m\lambda$, where m is an integer and λ is the wavelength, we might expect a bright ring with centre T. Similarly, if $2t = (m + \frac{1}{2})\lambda$, we might expect a dark ring.



FIG. 29.9. Newton's rings, formed by interference of yellow light between a convex lens and a flat glass plate.

When a ray is reflected from an optically denser medium, however, a phase change of 180° occurs in the wave, which is equivalent to its acquiring an extra path difference of $\lambda/2$ (see also p. 692). The truth of this statement can be seen by the presence of the dark spot at the centre, T, of the rings. At this point there is no geometrical path difference between the rays reflected from the lower surface of the lens and H, so that they should be in phase when they are brought to a focus and should form a bright spot. The dark spot means, therefore, that one of the rays suffers

a phase change of 180° . Taking the phase change into account, it follows that

$$2t = m\lambda \text{ for a dark ring} \quad . \quad . \quad . \quad (1)$$

and
$$2t = (m + \frac{1}{2})\lambda \text{ for a bright ring} \quad . \quad . \quad . \quad (2)$$

where m is an integer. Young verified the phase change by placing oil of sassafras between a crown and a flint glass lens. This liquid had a refractive index greater than that of crown glass and less than that of flint glass, so that light was reflected at an optically denser medium at each lens. A bright spot was then observed in the middle of the Newton's rings, showing that no net phase change had now occurred.

The grinding of a lens surface can be tested by observing the appearance of the Newton's rings formed between it and a flat glass plate when monochromatic light is used. If the rings are not perfectly circular, the grinding is imperfect. See Fig. 29.9.

Measurement of Wavelength by Newton's Rings

The radius r of a ring can be expressed in terms of the thickness, t , of the corresponding layer of air by simple geometry. Suppose TO is produced to D to meet the completed circular section of the lower surface PQ of the lens, PO being perpendicular to the diameter TD through T,

Fig. 29.10. Then, from the well-known theorem concerning the segments of chords in a circle, $TO \cdot OD = QO \cdot OP$. But $AT = r = PO$, $QO = OP = r$, $AP = t = TO$, and $OD = 2a - OT = 2a - t$.

$$\therefore t(2a - t) = r \times r = r^2$$

$$\therefore 2at - t^2 = r^2$$

But t^2 is very small compared with $2at$, as a is large.

$$\therefore 2at = r^2$$

$$\therefore 2t = \frac{r^2}{a} \quad \text{. (i)}$$

But $2t = (m + \frac{1}{2})\lambda$ for a bright ring.

$$\therefore \frac{r^2}{a} = (m + \frac{1}{2})\lambda . \quad . \quad . \quad . \quad . \quad (3)$$

The first bright ring obviously corresponds to the case of $m = 0$ in equation (3); the second bright ring corresponds to the case of $m = 1$. Thus the radius of the 15th bright ring is given from (3) by $r^2/a = 14\frac{1}{2}\lambda$, from which $\lambda = 2r^2/29a$. Knowing r and a , therefore, the wavelength λ can be calculated. Experiment shows that the rings become narrower when blue or violet light is used in place of red light, which proves, from equation (3), that the wavelength of violet light is shorter than the wavelength of red light. Similarly it can be proved that the wavelength of yellow light is shorter than that of red light and longer than the wavelength of violet light.

The radius r of a particular ring can be found by using a travelling microscope to measure its diameter. The radius of curvature, a , of the lower surface of the lens can be measured accurately by using light of known wavelength λ' , such as the green in a mercury-vapour lamp or the yellow of a sodium flame; since $a = r^2/(m + \frac{1}{2})\lambda'$ from (3), the radius of curvature a can be calculated from a knowledge of r , m , λ' .

Visibility of Newton's Rings

When white light is used in Newton's rings experiment the rings are coloured, generally with violet at the inner and red at the outer edge. This can be seen from the formula $r^2 = (m + \frac{1}{2}) \lambda a$, (3), as $r^2 \propto \lambda$. Newton gave the following list of colours from the centre outwards:

First order: Black, blue, white, yellow, orange, red. *Second order:* Violet, blue, green, yellow, orange, red. *Third order:* Purple, blue, green, yellow, orange, red. *Fourth order:* Green, red. *Fifth order:* Greenish-blue, red. *Sixth order:* Green-blue, pale-red. *Seventh order:* Greenish-blue, reddish-white. Beyond the seventh order the colours overlap and

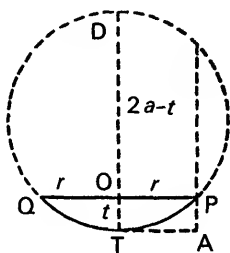


FIG. 29.10.
Theory of radius of
Newton's rings.

hence white light is obtained. The list is known generally as "Newton's scale of colours". Newton left a detailed description of the colours obtained with different thicknesses of air.

When Newton's rings are formed by sodium light, close examination shows that the clarity, or visibility, of the rings gradually diminishes as one moves outwards from the central spot, after which the visibility improves again. The variation in clarity is due to the fact that sodium light is not monochromatic but consists of *two* wavelengths, λ_2 , λ_1 , close to one another. These are (i) $\lambda_2 = 5890 \times 10^{-8}$ cm (D_2), (ii) $\lambda_1 = 5896 \times 10^{-8}$ cm (D_1). Each wavelength produces its own pattern of rings, and the ring patterns gradually separate as m , the number of the ring, increases. When $m\lambda_1 = (m + \frac{1}{2})\lambda_2$, the bright rings of one wavelength fall in the dark spaces of the other and the visibility is a minimum. In this case

$$5896m = 5890(m + \frac{1}{2}).$$

$$\therefore m = \frac{5890}{12} = 490 \text{ (approx.)}$$

At a further number of ring m_1 , when $m_1\lambda_1 = (m_1 + 1)\lambda_2$, the bright (and dark) rings of the two ring patterns coincide again, and the clarity, or visibility, of the interference pattern is restored. In this case

$$5896m_1 = 5890(m_1 + 1),$$

from which $m_1 = 980$ (approx.). Thus at about the 500th ring there is a minimum visibility, and at about the 1000th ring the visibility is a maximum.

It may be noted here that the bands in films of varying thickness, such as Newton's rings and the air-wedge bands, p. 692, appear to be formed in the film itself, and the eye must be focused on the film to see them. We say that the bands are "localised" at the film. With a thin film of uniform thickness, however, bands are formed by parallel rays which enter the eye, and these bands are therefore localised at infinity.

"Blooming" of Lenses

Whenever lenses are used, a small percentage of the incident light is reflected from each surface. In compound lens systems, as in telescopes and microscopes, this produces a background of unfocused

light, which results in a reduction in the clarity of the final image. There is also a reduction in the intensity of the image, since less light is transmitted through the lenses.

The amount of reflected light can be considerably reduced by evaporating a thin coating of a fluoride salt such as magnesium fluoride on to the surfaces, Fig. 29.11. Some of the light, of average

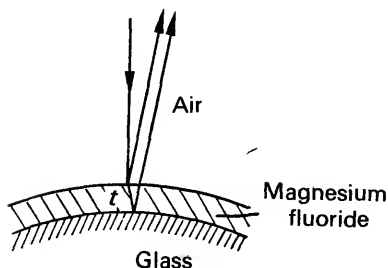


FIG. 29.11. Blooming of lens.

wavelength λ , is then reflected from the air-fluoride surface and the remainder penetrates the coating and is partially reflected from the fluoride-glass surface. Destructive interference occurs between the two reflected beams when there is a phase difference of 180° , or a path difference of $\lambda/2$, as the refractive index of the fluoride is less than that of glass. Thus if t is the required thickness of the coating and n' its refractive index, $2n't = \lambda/2$. Hence $t = \lambda/4n' = 6 \times 10^{-5}/(4 \times 1.38)$, assuming λ is 6×10^{-5} cm and n' is 1.38; thus $t = 1.1 \times 10^{-5}$ cm.

For best results n' should have a value equal to about \sqrt{n} , where n is the refractive index of the glass lens. The intensities of the two reflected beams are then equal, and hence complete interference occurs between them. No light is then reflected back from the lens. In practice, complete interference is not possible simultaneously for every wavelength of white light, and an average wavelength for λ , such as green-yellow, is chosen. "Bloomed" lenses effect a marked improvement in the clarity of the final image in optical instruments.

Lloyd's Mirror

In 1834 LLOYD obtained interference bands on a screen by using a plane mirror M , and illuminating it with light nearly at grazing incidence,

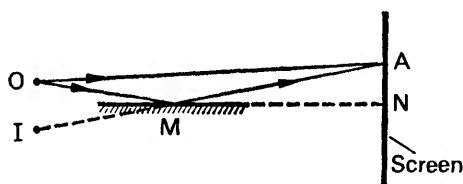


FIG. 29.12. Lloyd's mirror experiment.

coming from a slit O parallel to the mirror, Fig. 29.12. A point such as A on the screen is illuminated (i) by a ray OA and (ii) by a ray OM reflected along MA , which appears to come from the virtual image I of O in the mirror. Since O and I are close coherent sources interference bands are obtained on the screen.

Experiment showed that the band at N , which corresponds to the point of intersection of the mirror and the screen, was *dark*; since $ON = IN$, this band might have been expected, before the experiment was carried out, to be bright. Lloyd concluded that a phase change of 180° , equivalent to half a wavelength, occurred by reflection at the mirror surface, which is a denser surface than air (see p. 692).

Interference in Thin Films

The colours observed in a soap-bubble or a thin film of oil in the road are due to an interference phenomenon; they are also observed in thin transparent films of glass.

Consider a ray AO of monochromatic light incident on a thin parallel-sided film of thickness t and refractive index n . Fig. 29.13 is exaggerated for clarity. Some of the light is reflected at O along ON , while the

remainder is refracted into the film, where reflection occurs at B. The ray BC then emerges into the air along CM, which is parallel to ON. The incident ray AO thus divides at O into two beams of different amplitude which are coherent, and if ON, CM are combined by a lens, or by the eye-lens, a bright or dark band is observed according to the path difference of the rays.

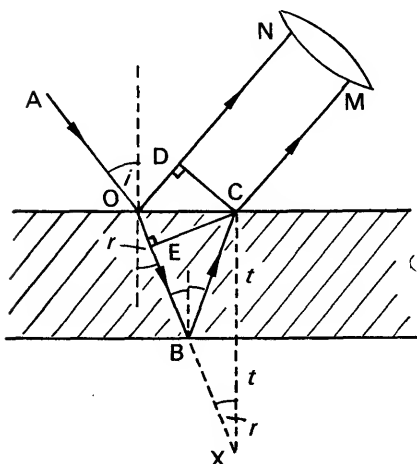


FIG. 29.13. Interference in thin films.

The time taken for light to travel a distance y in a medium of refractive index n is y/v , where v is the velocity of light in the medium. In this time, a distance $c \times y/v$ is travelled in air, where c is the velocity in air. But $n = c/v$. Hence the optical path of a length y in a medium of refractive index n is ny . The optical path difference between the two rays ON and OBCM is thus $n(OB + BC) - OD$, where CD is perpendicular to ON, Fig. 29.13. If CE is the perpendicular from C to OB, then $OD/OE = \sin i/\sin r = n$, so that $nOE = OD$.

$$\therefore \text{optical path difference} = n(EB + BC) = n(EB + BX) = n.EX.$$

$$= 2nt \cos r,$$

where r is the angle of refraction in the film. With a phase change of 180° by reflection at a denser medium, a bright band is therefore obtained when $2nt \cos r + \lambda/2 = m\lambda$,

$$\text{or} \quad 2nt \cos r = (m - \frac{1}{2})\lambda. \quad (i)$$

$$\text{For a dark band,} \quad 2nt \cos r = m\lambda. \quad (ii)$$

Colours in Thin Films

The colours in thin films of oil or glass are due to interference from an extended source such as the sky or a cloud. Fig. 29.14 illustrates interference between rays from points O_1, O_2 respectively on the extended source. Each ray is reflected and refracted at A_1, A_2 on the film, and enter the eye at E_1 . Although O_1, O_2 are non-coherent, the eye will see the same colour of a particular wavelength λ if $2nt \cos r = (m - \frac{1}{2})\lambda$. The separation of the two rays from A_1 or from A_2 must be less than the diameter of the eye-pupil for interference to occur, and this is the case only for thin films. The angle of refraction r is determined by the angle of incidence, or reflection, at the film. The particular colour seen thus depends on the position of the eye. At E_2 , for example, a different colour will be seen from another point O_3 on the extended source. The variation of θ and hence r is small when the eye observes a particular area of the film, and hence a band of a particular colour, such as $A_1 A_2$, is the con-

tour of paths of *equal inclination* to the film. The bands are localised at infinity, since the rays reaching E_1 or E_2 are parallel.

If a thin wedge-shaped film is illuminated by an extended source, as shown on p. 692 or in Newton's rings, the bands seen are contours of *equal thickness* of the film.

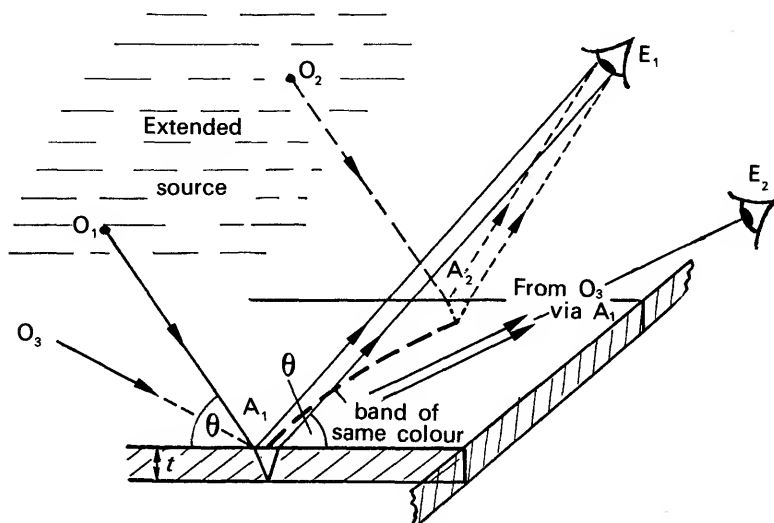


FIG. 29.14. Colours in thin films.

Vertical Soap Film Colours

An interesting experiment on thin films, due to C. V. Boys, can be performed by illuminating a vertical soap film with monochromatic light. At first the film appears uniformly coloured. As the soap drains to the bottom, however, a wedge-shaped film of liquid forms in the ring, the top of the film being thinner than the bottom. The thickness of the wedge is constant in a horizontal direction, and thus horizontal bright and dark bands are observed across the film. When the upper part of the film becomes extremely thin a black band is observed at the top (compare the dark central spot in Newton's rings experiment), and the film breaks shortly afterwards.

With white light, a succession of broad coloured bands is first observed in the soap film. Each band contains colours of the spectrum, red to violet. The bands widen as the film drains, and just before it breaks a black band is obtained at the top.

For normal incidence of white light, a particular wavelength λ is seen where the optical path difference due to the film $= (m - \frac{1}{2})\lambda$ and m is an integer. Thus a red colour of wavelength 7.0×10^{-5} cm is seen where the optical path difference is 3.5×10^{-5} cm, corresponding to $m = 1$. No other colour is seen at this part of the thin film. Suppose, however, that another part of the film is much thicker and the optical path difference here is $21 \times 3.5 \times 10^{-5}$ cm. Then a red colour of wavelength 7.0×10^{-5} cm, $m = 11$, an orange colour of wavelength about

6.4×10^{-5} cm, $m = 12$, a yellow wavelength about 5.9×10^{-5} cm, $m = 13$, and other colours of shorter wavelengths corresponding to higher integral values of m , are seen at the same part of the film. These colours all overlap and produce a white colour. If the film is thicker still, it can be seen that numerous wavelengths throughout the visible spectrum are obtained and the film then appears uniformly white.

Monochromatic light and Thin Parallel Films

If a thin parallel film is illuminated by a beam of monochromatic light, obtained by using an extended or broad source such as a bunsen burner sodium flame, a number of *circular* bright and dark curves can be seen. Fig. 29.15 illustrated how interference is obtained from the light originating from points a, b which is refracted at an angle α into the film. This is similar to Fig. 29.14 if B represents an eye-lens.

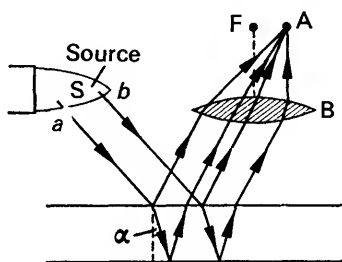


FIG. 29.15. Interference with extended source.

The emergent rays are combined by the eye-lens or a glass lens B, and a dark band is formed at A if $2nt \cos \alpha = m\lambda$, with the usual notation. If the light is incident on the film in every plane a circular band is obtained, whose centre is F, the focus of B. It is a band of 'equal inclination'.

When a *parallel* beam of monochromatic light is incident on the thin film, the angle of refraction r in the film and the thickness t are constant. The film thus appears uniformly bright at all points if the condition $2nt \cos r = (m + \frac{1}{2})\lambda$ is obeyed, and is uniformly dark if $2nt \cos r = m\lambda$. If the film is illuminated by a parallel beam of *white* light, the transmitted light appears to have dark bands across it when viewed through a spectroscope. The latter separates the colours, and a dark band is obtained where the condition $2nt \cos r = (m + \frac{1}{2})\lambda$ is satisfied for the particular wavelength, since we are now concerned with transmitted light.

EXAMPLE

What are Newton's rings and under what conditions can they be observed? Explain how they can be used to test the accuracy of grinding of the face of a lens. The face of a lens has a radius of curvature of 50 cm. It is placed in contact with a flat plate and Newton's rings are observed normally with reflected light of wavelength 5×10^{-5} cm. Calculate the radii of the fifth and tenth bright rings. (C.)

First parts. See text.

Second part. With the usual notation, for a bright ring we have

$$2t = (m + \frac{1}{2})\lambda, \quad \text{. (i)}$$

where t is the corresponding thickness of the layer of air.

But, from geometry, $2t = \frac{r^2}{a}$ (ii)

where r is the radius of the ring and a is the radius of curvature of the lens face (p. 695).

$$\therefore \frac{r^2}{a} = (m + \frac{1}{2})\lambda$$

$$\therefore r^2 = (m + \frac{1}{2})\lambda a \quad \dots \dots \dots (iii)$$

The first ring corresponds to $m = 0$ from equation (iii). Hence the fifth ring corresponds to $m = 4$, and its radius r is thus given by

$$\therefore r = \sqrt{\frac{9 \times 5 \times 10^{-5} \times 50}{2}} = 0.106 \text{ cm.}$$

The tenth ring corresponds to $m = 9$ in equation (iii), and its radius is thus given by

$$r^2 = 9\frac{1}{2} \times 5 \times 10^{-5} \times 50$$

$$\therefore r = 0.154 \text{ cm.}$$

DIFFRACTION OF LIGHT

In 1665 GRIMALDI observed that the shadow of a very thin wire in a beam of light was much broader than he expected. The experiment was repeated by Newton, but the true significance was only recognised more than a century later, after Huygens' wave theory of light had been resurrected. The experiment was one of a number which showed that light could bend round corners in certain circumstances.

We have seen how interference patterns, for example, bright and dark bands, can be obtained with the aid of two sources of light close to each other. These sources must be coherent sources, i.e., they must have the same amplitude and frequency, and always be in phase with each other. Consider two points on the *same wavefront*, for example the two points A, B, on a plane wavefront arriving at a narrow slit in a screen, Fig. 29.16. A and B can be considered as secondary sources of light, an aspect introduced by Huygens in his wave theory of light (p. 676); and as they are on the same wavefront, A and B have identical amplitudes

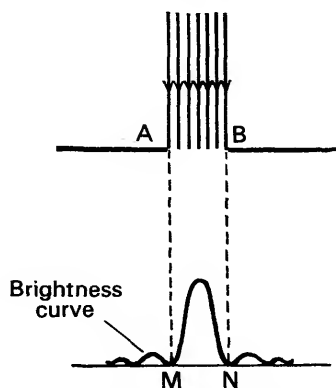


FIG. 29.16. Diffraction of light.



FIG. 29.17. Diffraction rings in the shadow of a small circular disc. The bright spot is at the centre of the geometrical shadow.

and frequencies and are in phase with each other. Consequently A, B, are coherent sources, and we can expect to find an interference pattern on a screen in front of the slit, provided the latter is small compared with the wavelength of light. For a short distance beyond the edges M, N, of the projection of AB, i.e., in the geometrical shadow, observation shows that there are some alternate bright and dark bands. See Fig. 29.20.

Thus light can travel round corners. The phenomenon is called *diffraction*, and it has enabled scientists to measure accurately the wavelength of light.

If a source of white light is observed through the eyelashes, a series of coloured images can be seen. These images are due to interference between sources on the same wavefront, and the phenomenon is thus an example of diffraction. Another example of diffraction was unwittingly deduced by POISSON at a time when the wave theory was new. Poisson considered mathematically the combined effect of the wavefronts round a circular disc illuminated by a distant small source of light, and he came to the conclusion that the light should be visible beyond the disc in the middle of the geometrical shadow. Poisson thought this was impossible; but experiment confirmed his deduction, and he became a supporter of the wave theory of light. See Fig. 29.17.

Diffraction at Single Slit

We now consider diffraction at a single slit in more detail. Suppose parallel light is incident on a narrow rectangular slit AB, Fig. 29.18.

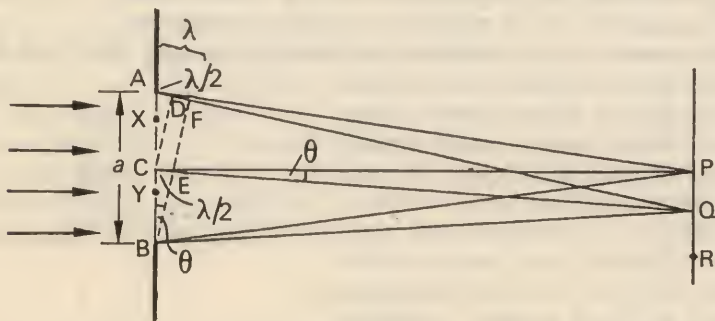


FIG. 29.18. Diffraction at single slit.

Each point on the same wavefront between A, B acts as a secondary centre of disturbance, and sends out wavelets beyond the slit. All the secondary centres are coherent, and their combined effect at any point such as P or Q can be found by summing the individual waves there,

from the Principle of Superposition. The mathematical treatment is beyond the scope of this book. The general effect, however, can be derived by considering the two halves AC, CB of the wavefront AB. At a point P equidistant from A and B, corresponding secondary centres in AC, CB respectively, such as X and Y, are also equidistant from P. Consequently wavelets arrive in phase at P. When AB is of the order of a few wavelengths of light the resultant amplitude at P due to the whole wavefront AB is therefore large, and thus a bright band is obtained at P.

As we move from P parallel to AB, points are obtained where the secondary wavelets from the two halves of the wavefront become more and more out of phase on arrival and the brightness thus diminishes. Consider a point Q, where AQ is half a wavelength longer than CQ. A disturbance from A, and one from C, then arrive at Q 180° out of phase. This is also practically the case for all corresponding points such as X, Y on the two halves of the wavefront. In particular, CQ and BQ differ practically by $\lambda/2$, where C is the extreme point in the upper half of the wavefront and B is the extreme point on the lower half. Thus Q corresponds to the edge or minimum intensity of the central band round P, Fig. 29.19. As we move farther away from Q parallel to AB, the

intensity rises again to a much smaller maximum at R, where $AR - BR = 3\lambda/2$, Fig. 29.18. To explain this, one can imagine the wavefront AB in Fig. 29.18 divided into three equal parts. Two parts annul each other's displacements at R as just explained, leaving one-third of the wavefront, which produces a much less bright band at R than at P. Calculation shows that the maximum intensity of the band at R is less than 5 per cent of that of the central band at P. Other subsidiary maxima and minima diffraction bands are obtained if the slit is very narrow. See Fig. 29.20.

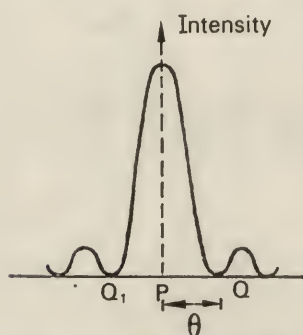


FIG. 29.19. Intensity variation – single slit.



FIG. 29.20. Diffraction bands formed by a single small rectangular aperture.

Width of Central Band. Rectilinear Propagation

The angular width of the central bright band is 2θ , where θ is the angular width between the maximum intensity direction P and the mini-

imum at Q, Fig. 29.19. From Fig. 29.18, it can be seen that the line CQ to the edge of the central band makes an angle θ with the direction CP of the incident light given by

$$\sin \theta = \frac{AF}{AB} = \frac{AD + CE}{AB} = \frac{\lambda/2 + \lambda/2}{a} = \frac{\lambda}{a},$$

where $a = AB$. When the slit is widened and a becomes large compared with λ , then $\sin \theta$ is very small and hence θ is very small. In this case the directions of the minimum and maximum intensities of the central band are very close to each other, and practically the whole of the light is confined to a direction immediately in front of the incident direction, that is, no spreading occurs. This explains the *rectilinear propagation of light*. When the slit width a is very small and equal to 2λ , for example, then $\sin \theta = \lambda/a = 1/2$, or $\theta = 30^\circ$. The light waves now spread round through 30° on either side of the slit.

These results are true for any wave phenomenon. In the case of an electromagnetic wave of 3 cm wavelength, a slit of these dimensions produces sideways spreading. Sound waves of a particular frequency 256 Hz have a wavelength of about 1.3 m. Consequently, sound waves spread round corners or apertures such as a doorway, which have comparable dimensions to their wavelengths.

Diffraction in Telescope Objective

When a parallel beam of light from a distant object such as a star S_1 enters a telescope objective L, the lens collects light through a circular opening and forms a diffraction pattern of the star round its principal focus, F. This is illustrated in the exaggerated diagram of Fig. 29.21.

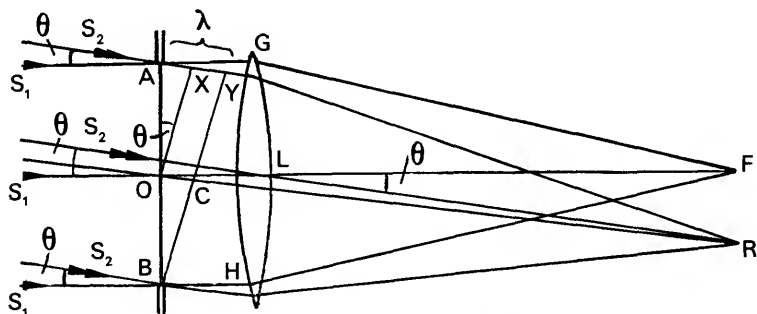


FIG. 29.21. Diffraction in telescope objective.

Consider an incident plane wavefront AB from the star S_1 , and suppose for a moment that the aperture is rectangular. The diffracted rays such as AG, BH normal to the wavefront are incident on the lens in a direction parallel to the principal axis LF. The optical paths AGF, BHF are equal. This is true for all other diffracted rays from points between A, B which are parallel to LF, since the optical paths to an image produced by a lens are equal. The central part F of the star pattern is therefore bright.

Now consider those diffracted rays from all points between AB which

enter the lens at an angle θ to the principal axis. This corresponds to a diffracted plane wavefront BY at an angle θ to AB . As described previously on p. 703, the wavefront AB can be divided into two halves, AO , OB . The rays from A and O in the two halves produce destructive interference if $AX = \lambda/2$, and likewise the extreme points O , B in the two halves produce destructive interference as $OC = \lambda/2$. Other corresponding points on the two halves also produce destructive interference. When the rays are collected and brought to a focus at R , darkness is thus obtained, that is, R is the edge of the central maximum of the star S_1 . As explained on p. 704, other subsidiary maxima may be formed round F .

The angle θ corresponding to the edge R is given by

$$\sin \theta = \frac{\lambda/2 + \lambda/2}{D} = \frac{\lambda}{D},$$

where D is the diameter of the lens aperture. This is the case where the opening can be divided into a number of rectangular slits. For a circular opening such as a lens (or the concave mirror of the Palomar telescope), the formula becomes $\sin \theta = 1.22\lambda/D$, and as θ is small, we may write $\theta = 1.22\lambda/D$.

Resolving Power

Suppose now that another distant star S_2 is at an angular distance θ from S_1 , Fig. 29.21. The maximum intensity of the central pattern of S_2 then falls on the minimum or edge of the central pattern of the star S_1 , corresponding to R in Fig. 29.22 (i). Experience shows that the two stars

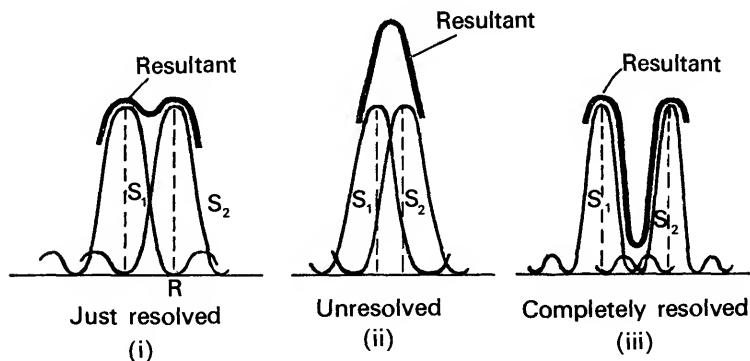


FIG. 29.22. Resolving power.

can then just be distinguished or *resolved*. Lord Rayleigh stated a criterion for the resolution of two objects, which is generally accepted: *Two objects are just resolved when the maximum intensity of the central pattern of one object falls on the first minimum or dark edge of the other.* Fig. 29.22 (i) shows the two stars just resolved. The resultant intensity in the middle dips to about 0.8 of the maximum, and the eye is apparently sensitive to the change here. Fig. 29.22 (ii) shows two stars S_1 , S_2 unresolved, and Fig. 29.22 (iii) the same stars completely resolved.

The angular distance θ between two distant stars just resolved is thus given by $\sin \theta = \theta = 1.22 \lambda/D$, where D is the diameter of the objective. This is an expression for the *limit of resolution*, or *resolving power*, of a telescope. The limit of resolution or resolving power increases when θ is smaller, as two stars closer together can then be resolved. Consequently telescope objectives of large diameter D give high resolving power. The Yerkes Observatory has a large telescope objective of about 100 cm. The angular distance θ between two stars which can just be resolved is thus given by

$$\theta = \frac{1.22 \lambda}{D} = \frac{1.22 \times 6 \times 10^{-5}}{100} = 7.3 \times 10^{-7} \text{ radians,}$$

assuming 6×10^{-5} cm for the wavelength of light. The Mount Palomar telescope has a parabolic mirror objective of aperture 5 metres, or 500 cm. The resolving power is thus five times as great as the Yerkes Observatory telescope. A large aperture D has also the advantage of collecting more light (p. 542). The Jodrell Bank radio telescope has a circular bowl of about 75 m, and for radio waves of 20 cm wavelength the resolving power, $\theta = 1.22 \lambda/D = 1.22 \times 20/7500$ radians $= 3 \times 10^{-3}$ radians (approx.).

Magnifying Power of Telescope and Resolving Power

If the width of the emergent beam from a telescope is greater than the diameter of the eye-pupil, rays from the outer edge of the objective do not enter the eye and hence the full diameter D of the objective is not used. If the width of the emergent beam is less than the diameter of the eye-pupil, the eye itself, which has a constant aperture, may not be able to resolve the distant objects. Theoretically, the angular resolving power of the eye is $1.22 \lambda/a$, where a is the diameter of the eye-pupil, but in practice an angle of 1 minute is resolved by the eye, which is more than the theoretical value.

Now the angular magnification, or magnifying power, of a telescope is the ratio α'/α , where α' is the angle subtended at the eye by the final image and α is the angle subtended at the objective (p. 533). To make the fullest use of the diameter D of the objective, the magnifying power should therefore be increased to the angular ratio given by

$$\frac{\text{resolving power of eye}}{\text{resolving power of objective}} = \frac{\pi/(180 \times 60)}{1.22 \times 6 \times 10^{-5}/D} = 4 D \text{ (approx.)}$$

In this case the telescope is said to be in "normal adjustment". Any further increase in magnifying power will make the distant objects appear larger, but there is no increase in definition or resolving power.

Brightness of Images in Telescope

In a telescope, the eye is placed at the *exit-pupil* or *eye-ring*, the circle through which the emergent beam passes (p. 532). The *entrance pupil* of the telescope is the aperture or diameter of the objective. If the area of the latter is A , then the smaller area of the exit-pupil is A/M^2 , where M is the angular magnification of the telescope in normal adjustment (see p. 535).

Consider a telescope used to observe (a) a small but finite area, or extended object, and (b) a point source, such as a star. Suppose that in each case the magnifying power is adjusted to make the exit pupil of the telescope equal to the eye pupil. In each case the luminous flux collected with the telescope is equal to the flux collected by an unaided eye multiplied by the ratio: area of objective/area of eye-pupil, which is M^2 .

Geometrically, the telescope magnifies the finite object area by a factor M^2 , but the image of the point object is still a point. Hence the area of the retinal image of the finite object is magnified by a factor M^2 when the telescope is used. On the other hand, since the eye-pupil is filled with light by the telescope, the retinal image of the point object with the telescope is the same as that without the telescope—it is the diffraction image for a point source. It will be seen that, for the finite object, the larger flux is spread over a larger image area, so that (apart from absorption losses in the telescope) the retinal illumination is unchanged. For the point object, however, the increased flux is spread over the same retinal area so that the brightness of the image is increased. On this account stars appear very much brighter when viewed by a telescope, whereas the brightness of the background, which acts as an extended object, remains about the same. Stars too faint to be seen with the naked eye become visible using a powerful telescope, and the number of stars seen thus increases considerably using a telescope.

Increasing Number of Slits. Diffraction Grating

On p. 703 we saw that the image of a single narrow rectangular slit is a bright central or principal maximum diffraction band, together with subsidiary maxima diffraction bands which are much less bright. Suppose that parallel light is incident on two more parallel close slits, and the light passing through the slits is received by a telescope focused at infinity. Since each slit produces a similar diffraction effect in the same direction, the observed diffraction pattern will have an intensity variation identical to that of a single slit. This time, however, the pattern is crossed by a number of interference bands, which are due to interference between slits (see *Young's experiment*, p. 688). The envelope of the intensity variation of the interference bands follow the diffraction pattern variation due to a single slit. In general, if I_s is the intensity at a point due to interference and I_d that due to diffraction, then the resultant intensity I is given by $I = I_d \times I_s$. Hence if $I_d = 0$ at any point, then $I = 0$ irrespective of the value of I_s .

As more parallel equidistant slits are introduced, the intensity and sharpness of the principal maxima increase and those of the subsidiary maxima decrease. The effect is illustrated roughly in Fig. 29.23. With several thousand lines per centimetre, only a few sharp principal maxima are seen in directions discussed shortly. Their angular separation depends only on the distance between successive slits. The slit width affects the intensity of the higher order principal maxima; the narrower the slit, the greater is the diffraction of light into the higher orders.

A *diffraction grating* is a large number of close parallel equidistant

slits, ruled on glass or metal; it provides a very valuable means of studying spectra. If the width of a slit or clear space is a and the thickness of a ruled opaque line is b , the spacing d of the slits is $(a + b)$. Thus with a grating of 6000 lines per centimetre, the spacing $d = 1/6000$ centimetre $= 17 \times 10^{-5}$ cm, or a few wavelengths of visible light.

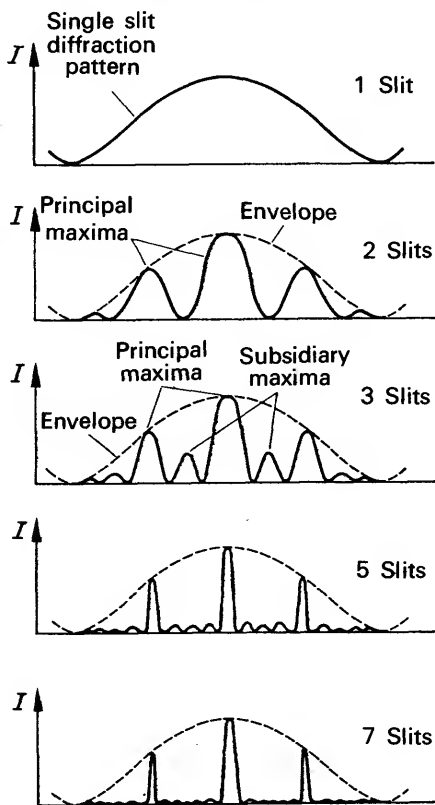


FIG. 29.23. Principal maxima with increasing slits.

Principal Maxima of Grating

The angular positions of the principal maxima produced by a diffraction grating can easily be found. Suppose X , Y are corresponding points in consecutive slits, where $XY = d$, and the grating is illuminated normally by monochromatic light of wavelength λ , Fig. 29.24. In a direction θ , the diffracted rays XL , YM have a path difference XA of $d \sin \theta$. The diffracted rays from all other corresponding points in the two slits have a path difference of $d \sin \theta$ in the same direction. Other pairs of slits throughout the grating can be treated in the same way. Hence bright or principal maxima are obtained when

$$d \sin \theta = m\lambda, \quad \dots \dots \dots (i)$$

where m is an integer, if all the diffracted parallel rays are collected by a telescope focused at infinity. The images corresponding to

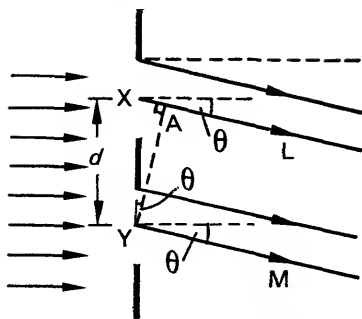


FIG. 29.24. Diffraction grating.

$m = 0, 1, 2, \dots$ are said to be respectively of the zero, first, second \dots orders respectively. The zero order image is the image where the path difference of diffracted rays is zero, and corresponds to that seen directly opposite the incident beam on the grating. It should again be noted that all points in the slits are secondary centres on the same wavefront and therefore coherent sources.

Diffraction Images

The *first order* diffraction image is obtained when $m = 1$. Thus

$$d \sin \theta = \lambda,$$

or

$$\sin \theta = \frac{\lambda}{d}.$$

If the grating has 6000 lines per centimetre (6000 cm^{-1}), the spacing of the slits, d , is $\frac{1}{6000}$ cm. Suppose yellow light, of wavelength

$\lambda = 5890 \times 10^{-8} \text{ cm}$, is used to illuminate the grating. Then

$$\sin \theta = \frac{\lambda}{d} = 5890 \times 10^{-8} \times 6000 = 0.3534$$

$$\therefore \theta = 20.7^\circ$$

The *second order* diffraction image is obtained when $m = 2$. In this case $d \sin \theta = 2\lambda$.

$$\therefore \sin \theta = \frac{2\lambda}{d} = 2 \times 5890 \times 10^{-8} \times 6000 = 0.7068$$

$$\therefore \theta = 45.0^\circ$$

If $m = 3$, $\sin \theta = 3\lambda/d = 1.060$. Since the sine of an angle cannot be greater than 1, it is impossible to obtain a third order image with this diffraction grating.

With a grating of 12000 lines per cm the diffraction images of sodium light would be given by $\sin \theta = m\lambda/d = m \times 5890 \times 10^{-8} \times 12000 = 0.7068 m$. Thus only $m = 1$ is possible here. As all the diffracted light is now concentrated in one image, instead of being distributed over several images, the first order image is very bright, which is an advantage.

Diffraction with Oblique Incidence

When a diffraction grating is illuminated by a monochromatic parallel beam PX, QY at an angle of incidence i , each point in the clear spaces acts as a secondary disturbance and diffracted beams emerge from the grating, Fig. 29.25. For a diffracted beam such as AB, making an angle of diffraction θ on the same side of the normal as PX or QY, the path difference between two typical rays PXA, QYB is $d(\sin i + \sin \theta)$. For a diffracted beam such as CD on the other side of the normal, the path difference between typical

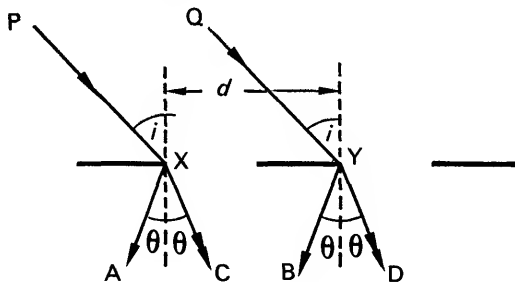


FIG. 29.25. Diffraction with oblique incidence.

rays PXC, QYD is $d(\sin i - \sin \theta)$. Thus, generally, a bright diffraction image is seen when $d(\sin i \pm \sin \theta) = m\lambda$, where m is an integer.

The zero order or central image is obtained in a direction opposite to the incident beam PX, QY. The first order diffraction image is obtained at angles θ on either side of this direction given respectively by $d(\sin i + \sin \theta) = \lambda$ and $d(\sin i - \sin \theta) = \lambda$. Diffraction images of higher order are obtained from similar formulae.

Reflection gratings can be used when light of particular wavelengths are absorbed by materials used in making transmission gratings. In this case the light is diffracted back into the incident medium at the clear spaces, and the diffraction images of various orders are given by $d(\sin i \pm \sin \theta) = m\lambda$.

Measurement of Wavelength

The wavelength of monochromatic light can be measured by a diffraction grating in conjunction with a spectrometer. The collimator C and telescope T of the instrument are first adjusted for parallel light (p. 445), and the grating P is then placed on the table so that its plane is perpendicular to two screws, Q, R, Fig. 29.26 (i). To level the table so that the plane of P is parallel to the axis of rotation of the telescope, the latter is first placed in the position T_1 directly opposite the illuminated slit of the collimator, and then rotated exactly through 90° to a position T_2 . The table is now turned until the slit is seen in T_2 by reflection at P, and one of the screws Q, R turned until the slit image is in the middle of the field of view. The plane of P is now parallel to the axis of rotation of the telescope. The table is then turned through 45° so that the plane of the grating is exactly perpendicular to the light from C, and the telescope is turned to a position T_3 to receive the first diffraction image, Fig. 29.26 (ii). If the lines of the grating are not parallel to the axis of rotation of the telescope, the image will not be in the middle of the field of view. The third screw is then adjusted until the image is central.

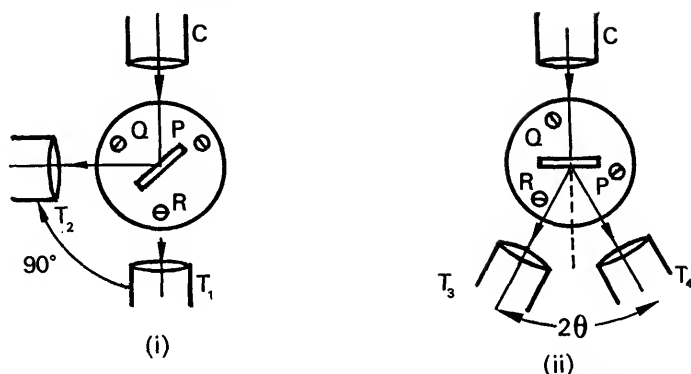


FIG. 29.26. Measurement of wavelength by diffraction grating.

The readings of the first diffraction image are observed on both sides of the normal. The angular difference is 2θ , and the wavelength is calculated from $\lambda = d \sin \theta$, where d is the spacing of the slits, obtained from the number of lines per centimetre of the grating. If a second order image is obtained for a diffraction angle θ , then $\lambda = d \sin \theta_1/2$.

Position of Image

If the grating lines are on the opposite side of the glass to the collimator C in Fig. 29.26 (ii), the light from C passes straight through the glass and the diffracted rays at the slits emerge into air. Suppose, however, that the grating is turned round so that the lines such as A, D face the collimator C, Fig. 29.27 (i). The rays are now diffracted into the glass and then refracted at B, F into the air at an angle θ to the normal. The optical path difference between the rays ABM, DFH from corresponding points A, D, is then

$$n.AB + BL - n.DF = BL$$

since $AB = DF$. But $BL = BF \sin \theta = d \sin \theta$. Consequently the angular

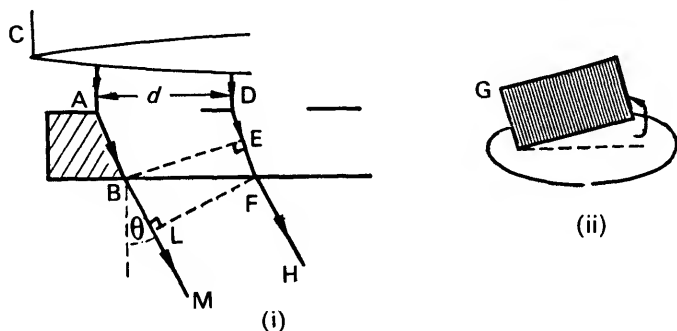


FIG. 29.27. Position of images with diffraction grating.

positions of the principal maxima diffraction images are given by $d \sin \theta = m\lambda$. Thus the images are observed at the same diffraction angles, no matter which side of the grating faces the collimator.

If the first order diffraction image is viewed in the telescope, and the grating G is turned round slightly in its own plane so that the lines are at a small angle to the vertical, the image of the slit moves round in the same direction, Fig. 29.27 (ii). The image then appears to move up or down in the field of view of the telescope, and disappears as the grating is turned round farther. The effect can be seen by viewing an electric lamp through a diffraction grating, and turning the grating in its own plane through 90° . The diffraction images of the lamp also rotate through 90° .

Spectra in Grating

If white light is incident normally on a diffraction grating, several coloured spectra are observed on either side of the normal, Fig. 29.28 (i). The first order diffraction images are given by $d \sin \theta = \lambda$, and as violet has a shorter wavelength than red, θ is less for violet than for red. Consequently the spectrum colours on either side of the incident white light are violet to red. In the case of a spectrum produced by dispersion in a glass prism, the colours range from red, the least deviated, to violet, Fig. 29.28 (ii). Second and higher order spectra are obtained with a diffraction grating on opposite sides of the normal, whereas only one spectrum is obtained with a glass prism. The angular spacing of the colours is also different in the grating and the prism.

If $d \sin \theta = m_1 \lambda_1 = m_2 \lambda_2$, where m_1, m_2 are integers, then a wavelength λ_1 in the m_1 order spectrum overlaps the wavelength λ_2 in the m_2 order. The

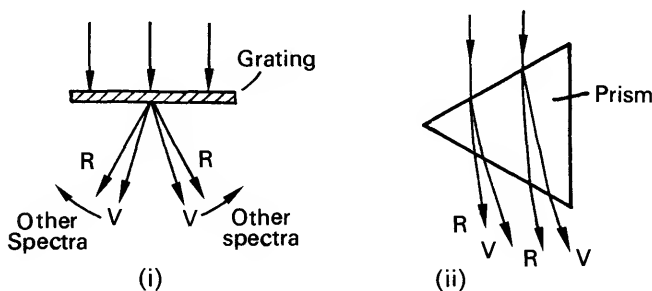


FIG. 29.28. Spectra in grating and prism.

extreme violet in the visible spectrum has a wavelength about 3.8×10^{-5} cm. The violet direction in the second order spectrum would thus correspond to $d \sin \theta = 2\lambda = 7.6 \times 10^{-5}$ cm, and this would not overlap the extreme colour, red, in the first order spectrum, which has a wavelength about 7.0×10^{-5} cm. In the second order spectrum, a wavelength λ_2 would be overlapped by a wavelength λ_3 in the third order if $2\lambda_2 = 3\lambda_3$. If $\lambda_2 = 6.9 \times 10^{-5}$ cm (red), then $\lambda_3 = 2\lambda_2/3 = 4.6 \times 10^{-5}$ cm (blue). Thus overlapping of colours occurs in spectra of higher orders than the first.

Dispersion by Grating

The *dispersion* of a grating, $d\theta/d\lambda$, is a measure of the change in angular position per unit wavelength change. Now $d \sin \theta = m\lambda$,

$$\begin{aligned}\therefore d \cos \theta \frac{d\theta}{d\lambda} &= m \\ \therefore \frac{d\theta}{d\lambda} &= \frac{m}{d \cos \theta} \quad \dots \quad (i)\end{aligned}$$

The dispersion thus increases with the order, m , of the image. It is also inversely proportional to the separation d of the slits, or, for a given grating width, directly proportional to the total number of slits on the grating. For a given order m , the dispersion increases when $\cos \theta$ is small, or when θ is large, which corresponds to the red wavelengths of the spectrum for normal incidence on the grating.

Resolving Power of Grating

For the m th order principal maximum of a grating, the path difference between diffracted rays from consecutive slits is $m\lambda$. The path difference AB between the extreme rays of the grating is thus $(N - 1)m\lambda$, where N is the total number of lines ruled on the grating, Fig. 29.29. The minimum intensity of the m th order principal maximum corresponds to a slightly different direction AC.

Now the discussion about the disturbances from various points across a wide slit (p. 702) can be applied to disturbances from various slits across a grating. It therefore follows that, for the minimum intensity, the path difference between disturbances from the first to the last slit is one wavelength, λ , more than that for the maximum intensity position. The path difference to the minimum is thus $(N - 1)m\lambda + \lambda$. The m th

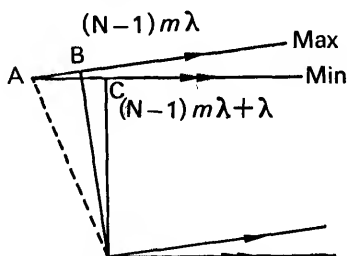


FIG. 29.29. Resolving power of diffraction grating.

order maximum of another wavelength λ' , differing slightly from λ , is formed by extreme rays which have a path difference of $(N - 1)m\lambda'$. From Rayleigh's criterion, the two wavelengths λ' and λ are just resolved when the maximum of λ' falls on the first minimum of λ . In this case,

$$\begin{aligned}(N - 1)m\lambda' &= (N - 1)m\lambda + \lambda \\ \therefore (N - 1)m(\lambda' - \lambda) &= \lambda \\ \therefore \frac{\lambda}{\lambda' - \lambda} &= (N - 1)m = Nm,\end{aligned}$$

since 1 is negligible compared with N .

$$\begin{aligned}\therefore \text{resolving power} &= Nm \\ \therefore \text{resolving power} &= \frac{Nd \sin \theta}{\lambda} \text{ or } \frac{Nd (\sin i \pm \sin \theta)}{\lambda},\end{aligned}$$

the former being the expression for light incident normally on the grating and the latter if the angle of incidence is i . Either expression shows that for a given angle of incidence and diffraction, it is the *total width Nd of the grating which determines its resolving power*. The number of rulings in that width affects the dispersion in a given order but has no effect on the resolving power in that order. Thus a grating of 5 cm width and 6000 lines per cm has twice the resolving power of a grating 2.5 cm wide which also has 6000 lines per cm. If a grating is only 0.3 cm wide it has only about 2000 lines on it of the same spacing, whereas a grating 5 cm wide would have 12000 lines.

The two sodium lines or doublet have wavelengths 5.890×10^{-5} and 5.896×10^{-5} cm respectively. The resolving power, R.P., required to distinguish them is given by:

$$\text{R.P.} = \frac{\lambda}{\lambda' - \lambda} = \frac{5.890 \times 10^{-5}}{0.006 \times 10^{-5}} = 1000 \text{ (approx.)}$$

Thus if a grating has 800 lines per cm, and the width covered by a telescope objective is 2.5 cm the sodium lines are clearly resolved in the first order images. If three-quarters of the grating is covered there are only 500 lines left, and the lines are now no longer resolved in the first order. They are just resolved in the second order images.

Resolving Power of Microscope

ABBE proposed a theory of image formation in a microscope which stated basically that if the structure of the illuminated object is regular (periodic), it acts like an illuminated diffraction grating. In this case the structure appears uniformly bright and unrecognisable if only the zero order image is collected by the microscope objective. If, in addition, the first order diffraction image is collected, the image plane in the objective contains alternate bright and dark strips or fringes in positions corresponding to images of the grating elements. The observer then recognises the grating structure, that is, the grating is "resolved". The more orders collected by the objective, the closer does the intensity distribution across the image plane resemble that transmitted by the object itself. The effect is analogous to the recognition of a note from a violin in Sound. This consists of a fundamental of the same frequency, together with overtones of higher frequency which gives the sound its timbre or quality. If only the fundamental is received, the note will not be recognisable as the note from the violin. The more overtones received in addition to the fundamental, the more faithful is the reproduction of the note.

An expression for the resolving power of a microscope can now be obtained. We require, for resolution, that a first order diffraction image is collected in addition to the zero order. Suppose that an object of regular structure is illuminated at an angle of incidence i by an oblique beam (Fig. 29.30). Then, if the first order image is just collected by the microscope objective,

$$d(\sin i \pm \sin \alpha) = \lambda,$$

where α is the half-angle subtended by the objective at the object O and

d is the grating spacing of the object. The minimum value of d occurs when $i = \alpha$ and $d(\sin i + \sin \alpha) = \lambda$.

$$\therefore 2d \sin \alpha = \lambda$$

$$\therefore d = \frac{\lambda}{2 \sin \alpha}$$

This expression for d gives the grating spacing of the finest regular structure of the object which can just be resolved. If a medium such as oil of refractive index n is used in the object space beneath the objective,

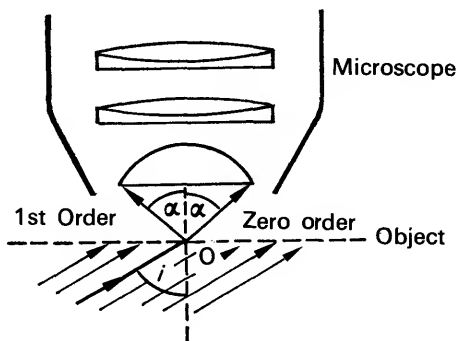


FIG. 29.30. Resolving power of microscope.

the least distance d or *limit of resolution* (also called the “*resolving power*”) is:

$$\text{limit of resolution} = \frac{\lambda}{2 n \sin \alpha}$$

The use of an oil-immersion objective was suggested by Abbe. The limit of resolution for the best optical microscopes is about 2×10^{-5} cm. The eye can resolve about 0.01 cm. The largest useful magnifying power of a microscope is one which magnifies the limit of resolution of the objective to that of the eye, and is about 1000 with glass lenses and visible light. Higher resolving powers may be obtained with ultra-violet light, from $\lambda/2 n \sin \alpha$. An *electron microscope*, which contains electron lenses and utilises electrons in place of light, has a limit of resolution less than 10^{-7} cm owing to the much shorter wavelength of moving electrons compared with that of light (p. 1077). Much larger useful magnifying powers, such as 100 000, are thus obtained by using electron microscopes in place of optical microscopes.

Wavelengths of Electromagnetic Waves

In this book we have encountered rays which affect the sensation of vision (*visible rays*), rays which cause heat (*infra-red rays*, p. 456), and rays which cause chemical action (ultra-violet rays, p. 456). As these rays are all due to electric and magnetic vibrations they are examples of *electromagnetic waves* (see p. 719). Scientists have measured the wavelengths of these waves by a diffraction grating method, and results show a gradual transition in the magnitudes of the wavelength from one type of ray to another. Thus infra-red rays have a longer wavelength

than visible rays, which in turn have a longer wavelength than ultra-violet rays. Radio waves are electromagnetic waves of longer wavelength

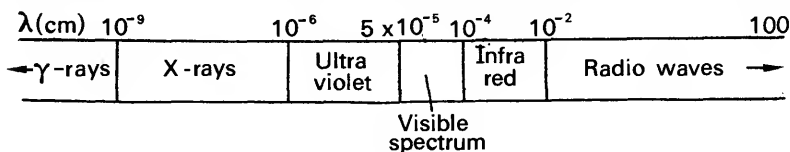


FIG. 29.31. Spectrum of electromagnetic waves (not to scale).

than infra-red rays, while X-rays and γ -rays are due to waves of shorter wavelength than ultra-violet waves. The whole spectrum of electromagnetic waves are shown in Fig. 29.31; this gives only an approximate value of the limits of the wavelength in the various parts of the spectrum, because these limits are themselves vague.

EXAMPLE

What is meant in optics by (a) interference, (b) diffraction? What part do each of these phenomena play in the production of spectra by a diffraction grating? A parallel beam of sodium light is incident normally on a diffraction grating. The angle between the two first order spectra on either side of the normal is $27^\circ 42'$. Assuming that the wavelength of the light is 5893×10^{-8} cm, find the number of rulings per cm on the grating. (*N.*)

First part. Briefly, interference is the name given to the phenomena obtained by the combined effect of light waves from two separate coherent sources; diffraction is the name given to the phenomena due to the combined effect of light waves from secondary sources on the same wavefront. In the diffraction grating, production of spectra is due to the interference between secondary sources on the same wavefront which are separated by a multiple of d , where d is the spacing of the grating rulings (p. 707).

Second part. The first order spectrum occurs at an angle $\theta = \frac{1}{2} \times 27^\circ 42' = 13^\circ 51'$.

But

$$d \sin \theta = \lambda$$

$$\therefore d = \frac{\lambda}{\sin \theta} = \frac{5893 \times 10^{-8}}{\sin 13^\circ 51'} \text{ cm}$$

$$\begin{aligned} \therefore \text{number of rulings per cm} &= \frac{1}{d} = \frac{\sin 13^\circ 51'}{5893 \times 10^{-8}} \\ &= 4062 \end{aligned}$$

POLARISATION OF LIGHT

We have shown that light is a wave-motion of some kind, i.e., that it is a travelling vibration. For a long time after the wave-theory was revived it was thought that the vibrations of light occurred in the same direction as the light wave travelled, analogous to sound waves. Thus light waves were thought to be longitudinal waves (p. 584). Observations and experiments, however, to be described shortly, showed that the vibrations of light occur in planes *perpendicular* to the direction along which the light wave travels, and thus light waves are *transverse* waves.

Polarisation of Transverse Waves

Suppose that a rope ABCD passes through two parallel slits, B, C, and is attached to a fixed point at D, Fig. 29.32 (i). Transverse waves can be set up along AB by holding the end A in the hand and moving it up and down in all directions perpendicular to AB, as illustrated by the arrows in the plane X. A wave then emerges along BC, but unlike the

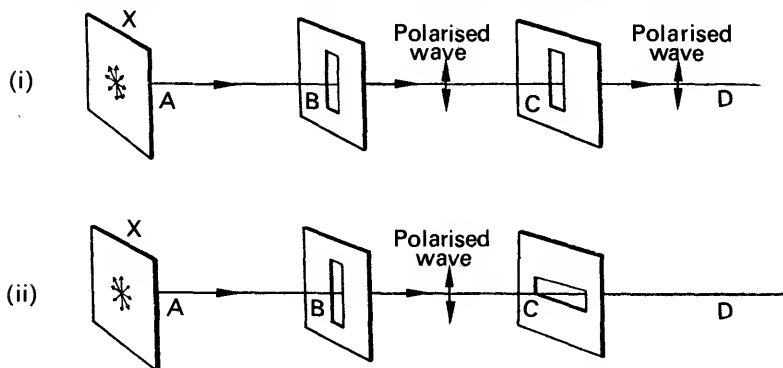


Fig. 29.32. Formation of plane-polarised waves.

waves along AB, which are due to transverse vibrations in every plane, it is due only to transverse vibrations parallel to the slit at B. This type of wave is called a *plane-polarised wave*. It shows a *lack of symmetry about the direction of propagation*, because a slit C allows the wave to pass through when it is parallel to B, but prevents it from passing when C is perpendicular to B, Fig. 29.32 (i), (ii). If B is turned so that it is perpendicular to the position shown in Fig. 29.32 (i), a polarised wave is again obtained along BC; but the vibrations which produce it are perpendicular to those shown between B and C in Fig. 29.32 (i).

Polarised Light

Years ago it was discovered accidentally that certain natural crystals affect light passing through them. *Tourmaline* is an example of such a crystal, *quartz* and *calcite* or *Iceland spar* are others (p. 720). Suppose two tourmaline crystals, P, Q, are placed with their axes, a , b , parallel, Fig. 29.33 (i). If a beam of light is incident on P, the light emerging from Q appears slightly darker. If Q is rotated slowly about the line of vision, with its plane parallel to P, the emergent light becomes darker and darker, and at one stage it disappears. In the latter case the axes a , b of the crystals are perpendicular, Fig. 29.33 (ii). When Q is rotated further the light reappears, and becomes brightest when the axes a , b are again parallel.

This simple experiment leads to the conclusion that light waves are *transverse waves*; otherwise the light emerging from Q could never be extinguished by simply rotating this crystal. The experiment, in fact, is analogous to that illustrated in Fig. 29.32, where transverse waves were set up along a rope and plane-polarised waves were obtained by means of a slit B. Tourmaline is a crystal which, because of its internal molecular

structure, transmits only those vibrations of light parallel to its axis. Consequently plane-polarised light is obtained beyond the crystal P, and no light emerges beyond Q when its axis is perpendicular to P. Fig. 29.33 should be compared with Fig. 29.32.

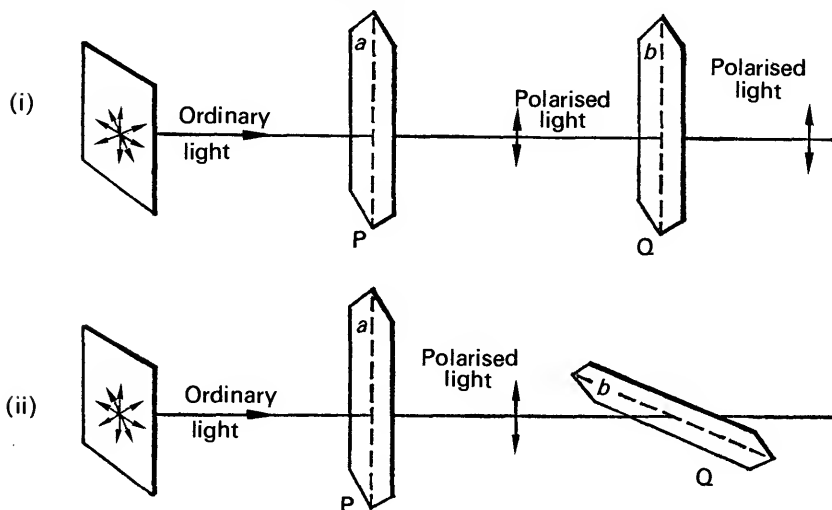


FIG. 29.33. Formation of plane-polarised light waves.

Vibrations in Unpolarised and Polarised Light

Fig. 29.34 (i) is an attempt to represent diagrammatically the vibrations of ordinary or unpolarised light at a point A when a ray travels in a direction AB. X is a plane perpendicular to AB, and ordinary (unpolarised) light may be imagined as due to vibrations which occur in

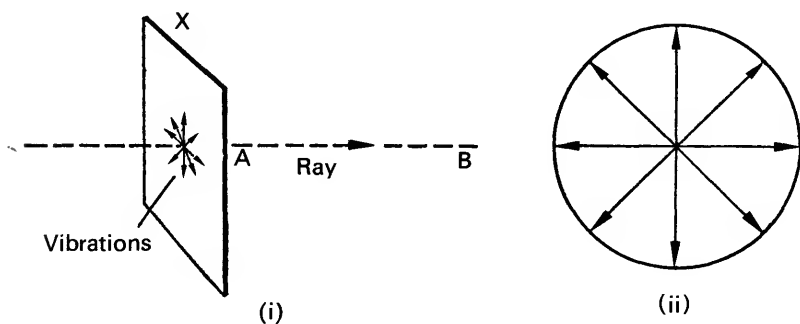


FIG. 29.34. (i). Vibrations occur in every plane perpendicular to AB.
(ii). Vibrations in ordinary light.

every one of the millions of planes which pass through AB and are perpendicular to X. As represented in Fig. 29.34 (ii), the amplitudes of the vibrations are all equal.

Consider the vibrations in ordinary light when it is incident on the tourmaline P in Fig. 29.33 (i). Each vibrations can be resolved into two components, one in a direction parallel to the axis a of the tourmaline P and the other in a direction m perpendicular to a , Fig. 29.35. Tourmaline absorbs the light due to the latter vibrations, known as the *ordinary* rays, allowing the light due to the former vibrations, known as the *extraordinary* rays, to pass through it. Thus plane-polarised light, due to the extraordinary rays, is produced by the tourmaline. Polaroid is a crystalline material, used in sun-glasses for example, which also has selective absorption.

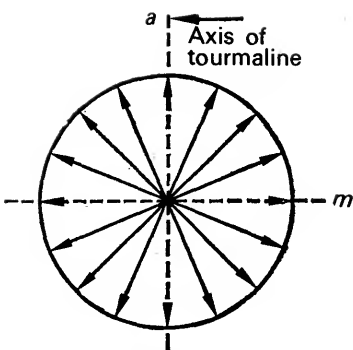


FIG. 29.35. Production of plane-polarised waves by tourmaline.

Light waves are electromagnetic waves. Theory and experiment show that the vibrations of light are *electromagnetic* in origin; a varying electric vector E is present, with a varying magnetic vector B which has the same frequency and phase. E and B are perpendicular to each other, and are in a plane at right angles to the ray of light, Fig. 29.36. Experiments have shown that the electric force in a light wave affects a photographic plate and causes fluorescence, while the magnetic force, though present, plays no part in this effect of a light wave. On this account the vibrations of the electric force, E , are now chosen as the “vibrations of light”, and the planes containing the vibrations shown in Fig. 29.35 (i), (ii) are those in which only the electric forces are present.

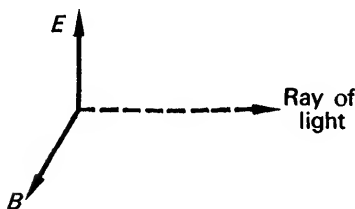


FIG. 29.36. Electromagnetic wave.

Polarised Light by Reflection

The production of polarised light by tourmaline is due to *selective absorption* of the “ordinary” rays. In 1808 MALUS discovered that polarised light is obtained when ordinary light is reflected by a plane sheet of glass (p. 719). The most suitable angle of incidence is about 56° , Fig. 29.37. If the reflected light is viewed through a tourmaline crystal which is slowly rotated about the line of vision, the light is practically extinguished at one position of the crystal. This proves that the light reflected by the glass is plane-polarised. Malus also showed that the light reflected by water is plane-polarised.

The production of the polarised light by the glass is explained as follows. Each of the vibrations of the incident (ordinary) light can be resolved into a component parallel to the glass surface and a component perpendicular to the surface. The light due to the components

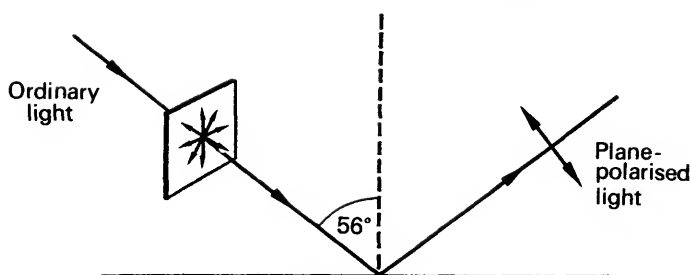


FIG. 29.37. Plane-polarised waves by reflection.

parallel to the glass is reflected, but the remainder of the light, due to the components perpendicular to the glass, is *refracted* into the glass. Thus the light reflected by the glass is plane-polarised.

Brewster's Law. Polarisation by Pile of Plates

The particular angle of incidence i on a transparent medium when the reflected light is almost completely plane-polarized is called the *polarising angle*. BREWSTER found that, in this case, $\tan i = n$, where n is the refractive index of the medium (*Brewster's law*). Since $\sin i / \sin r$, where r is the angle of refraction, it then follows that $\cos i = \sin r$, or $i + r = 90^\circ$. Thus the reflected and refracted beams are at 90° to each other.

The refracted beam contains light mainly due to vibrations perpendicular to that reflected and is therefore partially plane-polarised. Since refraction and reflection occur at both sides of a glass plate, the transmitted beam contains a fair percentage of plane-polarised light. A *pile of plates* increases the percentage, and thus provides a simple method of producing plane-polarised light. They are mounted inclined in a tube so that the ordinary (unpolarised) light is incident at the polarising angle, and the transmitted light is then fairly plane-polarised.

Polarisation by Double Refraction

We have already considered two methods of producing polarised light. The first observation of polarised light, however, was made by BARTHO-LINUS in 1669, who placed a crystal of Iceland spar on some words on a sheet of paper. To his surprise, two images were seen through the crystal. Bartholinus therefore gave the name of *double refraction* to the phenomenon, and experiments more than a century later showed that the crystal produced plane-polarised light when ordinary light was incident on it. See Fig. 29.38.

Iceland spar is a crystalline form of calcite (calcium carbonate) which cleaves in the form of a "rhomboid" when it is lightly tapped; this is a solid whose opposite faces are parallelograms. When a beam of unpolarised light is incident on one face of the crystal, its internal molecular structure produces two beams of polarised light, E, O, whose vibrations are perpendicular to each other, Fig. 29.39. If the incident direction AB is parallel to a plane known as the "principal section" of the crystal, one



FIG. 29.38. DOUBLE REFRACTION. A ring with a spot in the centre, photographed through a crystal of Iceland spar. The light forms two rings and two spots.

beam O emerges parallel to AB, while the other beam E emerges displaced in a different direction. As the crystal is rotated about the line of vision the beam E revolves round O. On account of this abnormal behaviour the rays in E are called “extraordinary” rays; the rays in O are known as “ordinary” rays (p. 719). Thus two images of a word on a paper, for example, are seen when an Iceland spar crystal is placed on top of it; one image is due to the ordinary rays, while the other is due to the extraordinary rays.

With the aid of an Iceland spar crystal Malus discovered the polarisation of light by reflection (p. 719). While on a visit to Paris he gazed

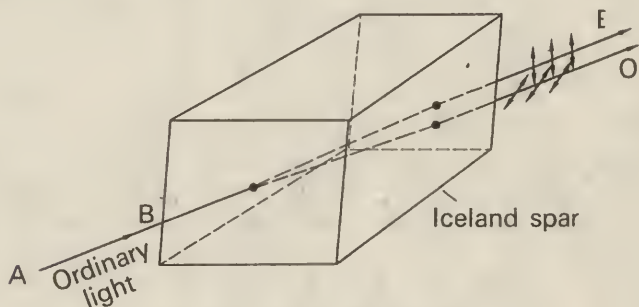


FIG. 29.39. Action of Iceland spar.

through the crystal at the light of the sun reflected from the windows of the Palace of Luxemburg, and observed that only one image was obtained for a particular position of the crystal when it was rotated slowly. The light reflected from the windows could not therefore be ordinary (unpolarised) light, and Malus found it was plane-polarised.

Nicol Prism

We have seen that a tourmaline crystal produces polarised light, and that the crystal can be used to detect such light (p. 717). NICOL designed a crystal of Iceland spar which is widely used for producing and detecting

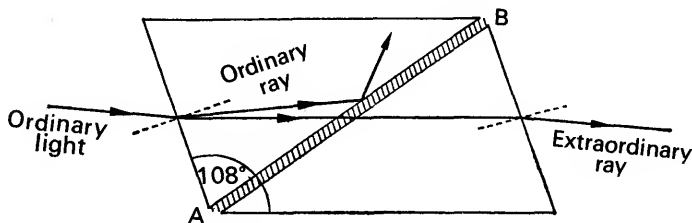


FIG. 29.40. Action of Nicol prism.

polarised light, and it is known as a *Nicol prism*. A crystal whose faces contain angles of 72° and 108° is broken into two halves along the diagonal AB, and the halves are cemented together by a layer of Canada balsam, Fig 29. 40. The refractive index of the crystal for the ordinary rays is 1.66, and is 1.49 for the extraordinary rays; the refractive index of the Canada balsam is about 1.55 for both rays, since Canada balsam does not produce polarised light. A critical angle thus exists between the crystal and Canada balsam for the ordinary rays, but not for the extraordinary rays. Hence total reflection of the former rays takes place at the canada balsam if the angle of incidence is large enough, as it is with the Nicol prism. The emergent light is then due to the extraordinary rays, and is polarised.

The prism is used like a tourmaline crystal to detect plane-polarised light, namely, the prism is held in front of the beam of light and is rotated. If the beam is plane-polarised the light seen through the Nicol prism varies in intensity, and is extinguished at one position of the prism.

Differences Between Light and Sound Waves

We are now in a position to distinguish fully between light and sound waves. The physical difference, of course, is that light waves are due to varying electric and magnetic forces, while sound waves are due to vibrating layers or particles of the medium concerned. Light can travel through a vacuum, but sound cannot travel through a vacuum. Another very important difference is that the vibrations of the particles in sound waves are in the same direction as that along which the sound travels, whereas the vibrations in light waves are perpendicular to the direction along which the light travels. Sound waves are therefore *longitudinal* waves, whereas light waves are *transverse* waves. As we have seen, sound waves can be reflected and refracted, and can give rise to interference phenomena; but no polarisation phenomena can be obtained with sound waves since they are longitudinal waves, unlike the case of light waves.

EXERCISES 29

Interference

1. Describe how to set up apparatus to observe and make measurements on the interference fringes produced by Young's slits. Explain how (i) the wavelengths of two monochromatic light sources could be compared, (ii) the separation of the slits could be deduced using a source of known wavelength. Establish any formula required.

State, giving reasons, what you would expect to observe (a) if a white light source were substituted for a monochromatic source, (b) if the source slit were then displaced slightly at right angles to its length in the plane parallel to the plane of the Young's slits. (L.)

2. Explain the formation of interference fringes by an air wedge and describe how the necessary apparatus may be arranged to demonstrate them.

Fringes are formed when light is reflected between the flat top of a crystal resting on a fixed base and a sloping glass plate. The lower end of the plate rests on the crystal and the upper end on a fixed knife-edge. When the temperature of the crystal is raised the fringe separation changes from 0.96 mm to 1.00 mm. If the length of the glass plate from knife-edge to crystal is 5.00 cm, and the light of wavelength 6.00×10^{-5} cm is incident normally on the wedge, calculate the expansion of the crystal. (L.)

3. Describe in detail how the radius of curvature of the spherical face of a planoconvex lens may be found by observations made on Newton's rings.

Two plane glass plates which are in contact at one edge are separated by a piece of metal foil 12.50 cm from that edge. Interference fringes parallel to the line of contact are observed in reflected light of wavelength 5460 Å and are found to be 1.50 mm apart. Find the thickness of the foil. (L.)

4. Describe, with the aid of a labelled diagram, how the wavelength of monochromatic light may be found using Young's slits. Give the theory of the experiment.

State, and give physical reasons for the features which are common to this method and to *either* the method based on Lloyd's mirror *or* that based on Fresnel's biprism.

In an experiment using Young's slits the distance between the centre of the interference pattern and the tenth bright fringe on either side is 3.44 cm and the distance between the slits and the screen is 2.00 m. If the wavelength of the light used is 5.89×10^{-7} m determine the slit separation. (N.)

5. Explain what is meant by the term *path-difference* with reference to the interference of two wave-motions.

Why is it not possible to see interference where the light beams from the headlamps of a car overlap?

Interference fringes were produced by the Young's slits method, the wavelength of the light being 6×10^{-5} cm. When a film of material 3.6×10^{-3} cm thick was placed over *one* of the slits, the fringe pattern was displaced by a distance equal to 30 times that between two adjacent fringes. Calculate the refractive index of the material. To which side are the fringes displaced?

(When a layer of transparent material whose refractive index is n and whose thickness is d is placed in the path of a beam of light, it introduces a path difference equal to $(n-1)d$.) (O. & C.)

6. Show how, with the aid of Huygens' idea of secondary wavelets, the wave theory of light will account for the laws of refraction and of reflexion at a plane surface.

Describe briefly Young's two-slit experiment and explain how it confirms the wave nature of light. (L.)

7. Describe, giving both theory and experimental detail, how you would find the radius of curvature of one surface of a convex lens by means of Newton's rings. You may assume that monochromatic light of a known wavelength is available.

Newton's rings are formed by reflexion between an equiconvex lens of focal length 100 cm made of glass of refractive index 1.50 and in contact with a plane glass plate of refractive index 1.60. Find the radius of the 5th bright ring using monochromatic light of wavelength 6000 Å.

Explain the changes which occur when oil of refractive index 1.55 fills the space between the lens and plate. ($1 \text{ Å} = 10^{-8} \text{ cm.}$) (N.)

8. Define *velocity*, *frequency* and *wavelength* for any wave motion, and deduce a relation between them. What do you understand by 'interference between waves' and 'coherent wave trains'? Explain why interference is not observed between the beams of two electric torches.

Deduce the relation connecting the refractive index of a material with the velocities of light in vacuo and in the material. State clearly the assumptions you make about wave fronts in order to do this.

Light passes through a single crystal of ruby 10.0 cm long and emerges with a wavelength of $6.94 \times 10^{-5} \text{ cm}$. If the critical angle of ruby for light of this wavelength is $34^\circ 50'$, calculate the number of wavelengths inside the crystal. (C.)

9. State the conditions necessary for the production of interference effects by two overlapping beams of light.

Describe fully *one* method for the production of interference fringes using light from a given monochromatic source. Show how with the aid of suitable measurements the wavelength of light emitted by the source may be determined with your apparatus.

Describe how the fringes produced by your apparatus would appear if a source of white light were employed instead of a monochromatic one. (O. & C.)

10. Explain how Newton's rings are formed, and describe how you would demonstrate them experimentally. How is it possible to predict the appearance of the centre of the ring pattern when (a) the surfaces are touching, and (b) the surfaces are not touching?

In a Newton's rings experiment one surface was fixed and the other movable along the axis of the system. As the latter surface was moved the rings appeared to contract and the centre of the pattern, initially at its darkest, became alternately bright and dark, passing through 26 bright phases and finishing at its darkest again. If the wavelength of the light was 5461 Å, how far was the surface moved and did it approach, or recede from, the fixed surface? Suggest one possible application of this experiment. (O. & C.)

11. Explain the formation of Newton's rings and describe how you would use them to measure the radius of curvature of the convex surface of a long-focus planoconvex lens.

The diameters of the m th and $(m + 10)$ th bright rings formed by such a lens resting on a plane glass surface are respectively 0.14 cm and 0.86 cm. When the space between lens and glass is filled with water the diameters of the q th and $(q + 10)$ th bright rings are respectively 0.23 cm and 0.77 cm. What is the refractive index of water? (L.)

12. What are the conditions essential for the production of optical interference fringes?

Explain how these conditions are satisfied in the case of (a) Young's fringes, and (b) thin film interference fringes. (N.)

13. Describe, in detail how you would arrange apparatus to observe, in monochromatic light, interference fringes formed by light reflected from two glass plates enclosing an air wedge. Show how the angle of the wedge could be obtained from measurements on the fringes.

Newton's rings are formed with light of wavelength 5.89×10^{-5} cm between the curved surface of a planoconvex lens and a flat glass plate, in perfect contact. Find the radius of the 20th dark ring from the centre if the radius of curvature of the lens surface is 100 cm. How will this ring move and what will its radius become if the lens and the plate are slowly separated to a distance apart of 5.00×10^{-4} cm? (L.)

14. What are the necessary conditions for interference of light to be observable? Describe with the aid of a labelled diagram how optical interference may be demonstrated using Young's slits. Indicate suitable values for all the distances shown.

How are the colours observed in thin films explained in terms of the wave nature of light? Why does a small oil patch on the road often show approximately circular coloured rings? (L.)

Diffraction

15. Describe and give the theory of an experiment to compare the wavelengths of yellow light from a sodium and red light from a cadmium discharge lamp, using a diffraction grating. Derive the required formula from first principles.

White light is reflected normally from a soap film of refractive index 1.33 and then directed upon the slit of a spectrometer employing a diffraction grating at normal incidence. In the first-order spectrum a dark band is observed with minimum intensity at an angle of $18^\circ 0'$ to the normal. If the grating has 5000 lines per cm, determine the thickness of the soap film assuming this to be the minimum value consistent with the observations. (L.)

16. Describe the phenomena which occur when plane waves pass (a) through a wide aperture, (b) through an aperture whose width is comparable with the wavelength of the waves.

How does the wave theory of light account for the apparent rectilinear propagation of light?

A diffraction grating has 6000 lines per cm. Calculate the angular separation between wavelengths 5.896×10^{-5} cm and 5.461×10^{-5} cm respectively after transmission through it at normal incidence, in the first-order spectrum. (O. & C.)

17. Describe two experiments to show the diffraction of light.

Describe how a diffraction grating may be used to measure the wavelength of sodium light, deriving any formulae employed. (L.)

18. What are the advantages and disadvantages of a diffraction grating as compared with a prism for the study of spectra?

A rectangular piece of glass 2 cm \times 3 cm has 18000 evenly spaced lines ruled across its whole surface, parallel to the shorter side, to form a diffraction grating. Parallel rays of light of wavelength 5×10^{-5} cm fall normally on the grating. What is the highest order of spectrum in the transmitted light?

What is the minimum diameter of a camera lens which can accept all the light of this wavelength in this order which leaves the grating on one side of the normal? (*O. & C.*)

19. In an experiment using a spectrometer in normal adjustment fitted with a plane transmission grating and using monochromatic light of wavelength 5.89×10^{-5} cm, diffraction maxima are obtained with telescope settings of $153^\circ 44'$, $124^\circ 5'$, $76^\circ 55'$ and $47^\circ 16'$, the central maximum being at $100^\circ 30'$. Show that these observations are consistent with normal incidence and calculate the number of rulings per cm of the grating.

If this grating is replaced by an opaque plate having a single vertical slit 2.00×10^{-2} cm wide, describe and explain the diffraction pattern which may now be observed. Contrast the appearance of this pattern with that produced by the grating. (*N.*)

20. (a) What is meant by (i) *diffraction*, (ii) *superposition* of waves? Describe one phenomenon to illustrate each in the case of *sound waves*.

(b) The floats of two men fishing in a lake from boats are 22.5 metres apart. A disturbance at a point in line with the floats sends out a train of waves along the surface of the water, so that the floats bob up and down 20 times per minute. A man in a third boat observes that when the float of one of his colleagues is on the crest of a wave that of the other is in a trough, and that there is then one crest between them. What is the velocity of the waves? (*O. & C.*)

21. Give an account of the theory of the production of a spectrum by means of a plane diffraction grating. How does it differ from the spectrum produced by means of a prism?

Parallel light consisting of two monochromatic radiations of wavelengths 6×10^{-5} cm and 4×10^{-5} cm falls normally on a plane transmission grating ruled with 5000 lines per cm. What is the angular separation of the second-order spectra of the two wavelengths? (*C.*)

22. A pure spectrum is one in which there is no overlapping of light of different wavelengths. Describe how you would set up a diffraction grating to display on a screen as close an approximation as possible to a pure spectrum. Explain the purpose of each optical component which you would use.

A grating spectrometer is used at normal incidence to observe the light from a sodium flame. A strong yellow line is seen in the first order when the telescope axis is at an angle of $16^\circ 26'$ to the normal to the grating. What is the highest order in which the line can be seen?

The grating has 4800 lines per cm; calculate the wavelength of the yellow radiation.

What would you expect to observe in the spectrometer set to observe the first-order spectrum if a small but very bright source of white light is placed close to the sodium flame so that the flame is between it and the spectrometer? (*O. & C.*)

23. Describe how you would determine the wavelength of monochromatic light using a diffraction grating and a spectrometer. Give the theory of the method.

A filter which transmits only light between 6300 Å and 6000 Å is placed between a source of white light and the slit of a spectrometer; the grating has 5000 lines to the centimetre; and the telescope has an objective of focal length 15 cm with an eyepiece of focal length 3 cm. Find the width in millimetres of the first-order spectrum formed in the focal plane of the objective. Find also the angular width of this spectrum seen through the eyepiece. (*O.*)

Polarisation

24. What is meant by *plane of polarisation*? Explain why the phenomenon of polarisation is met with in dealing with light waves, but not with sound waves.

Describe and explain the action of (a) a nicol prism, (b) a sheet of Polaroid.

How can a pair of Polaroid sheets and a source of natural light be used to produce a beam of light the intensity of which may be varied in a calculable manner? (L.)

25. Explain what is meant by the statement that a beam of light is *plane polarised*. Describe *one* experiment in each instance to demonstrate (a) polarisation by reflexion, (b) polarisation by double refraction, (c) polarisation by scattering.

The refractive index of diamond for sodium light is 2.417. Find the angle of incidence for which the light reflected from diamond is completely plane polarised. (L.)

26. Give an account of the action of (a) a single glass plate. (b) a Nicol prism, in producing plane-polarised light. State *one* disadvantage of *each* method.

Mention *two* practical uses of polarising devices. (N.)

27. Describe how, using a long, heavy rope, you would demonstrate (a) a plane-polarised wave, and (b) a stationary wave.

Give a short account, with diagrams, of *three* ways in which plane-polarised light is obtained (other than by using 'polaroid'). State some uses of polarised light.

Two polaroid sheets are placed close together in front of a lamp so that no light passes through them. Describe and explain what happens when one sheet is slowly rotated, the other remaining in its original position. (C.)

28. Answer *two* of the following:

(i) How may it be shown that the radiation from (a) a sodium lamp, and (b) a radio transmitter (such as a broadcasting station or a microwave source) consists of waves?

(ii) Explain what is meant by the polarisation of light, and describe how you would demonstrate it. Why is light from most light sources unpolarised?

(iii) When a diffraction grating is illuminated normally by monochromatic light an appreciable amount of light leaves the grating in certain directions. Explain this phenomenon, and show how these directions may be predicted. (O. & C.)

29. What is meant by (a) polarised light, (b) polarising angle? Describe and explain two methods for producing plane-polarised light.

Calculate the polarising angle for light travelling from water, of refractive index 1.33, to glass, of refractive index 1.53. (L.)

30. A beam of plane-polarised light falls normally on a sheet of Polaroid, which is at first set so that the intensity of the transmitted light, as estimated by a photographer's light-meter, is a maximum. (The meter is suitably shielded from all other illumination.) Describe and explain the way in which you would expect the light-meter readings to vary as the Polaroid is rotated in stages through 180° about an axis at right angles to its plane.

How would you show experimentally (a) that calcite is doubly refracting, (b) that the two refracted beams are plane polarised, in planes at right angles to one another, and (c) that in general the two beams travel through the crystal with different velocities? (O.)

31. What is plane-polarised light?

Explain why two images of an object are seen through a crystal of Iceland Spar. What would be seen if the object were viewed through two crystals, one of which was slowly rotated about the line of vision?

How would you produce a plane-polarised beam of light by reflection from a glass surface? (C.)

32. What is meant by the polarisation of light? How is polarisation explained on the hypothesis that light has wave properties?

Describe how polarisation can be produced and detected by reflexion. Mention another way of obtaining polarised light and describe how you would determine which of the two methods is the more effective.

Describe briefly *two* uses of polarised light. (N.)

33. Give an account of the evidence for believing that light is a wave motion. What reason is there for believing light waves to be transverse waves?

Two dishes *A* and *B* each contain liquid to a depth of 3.000 cm. *A* contains alcohol, *B* a layer of water on which is a layer of transparent oil. The depths of the oil and water are adjusted so that for monochromatic light passing vertically through them, the number of wavelengths is the same in *A* and *B*. Find the depth of the water layer, if the refractive indices of alcohol, water and oil are respectively 1.363, 1.333 and 1.475. (L.)

PART FOUR

Electricity and Atomic Physics

chapter thirty

Electrostatics

GENERAL PHENOMENA

IF a rod of ebonite is rubbed with fur, or a fountain-pen with a coat-sleeve, it gains the power to attract light bodies, such as pieces of paper or tin-foil or a suspended pith-ball. The discovery that a body could be made attractive by rubbing is attributed to Thales (640–548 B.C.). He seems to have been led to it through the Greeks' practice of spinning silk with an amber spindle; the rubbing of the spindle in its bearings caused the silk to adhere to it. The Greek word for amber is *elektron*, and a body made attractive by rubbing is said to be 'electrified'. This branch of Electricity, the earliest discovered, is called *Electrostatics*.

Conductors and Insulators

Little progress was made in the study of electrification until the sixteenth century A.D. Then Gilbert (1540–1603), who was physician-in-ordinary to Queen Elizabeth, found that other substances besides amber could be electrified: for example, glass when rubbed with silk. He failed to electrify metals, however, and concluded that to do so was impossible.

More than 100 years later—in 1734—he was shown to be wrong, by du Fay; du Fay found that a metal could be electrified by rubbing with fur or silk, but only if it were held in a handle of glass or amber; it could not be electrified if it were held directly in the hand. His experiments followed the discovery, by Gray in 1729, that electric charges could be transmitted through the human body, water, and metals. These are examples of *conductors*; glass and amber are examples of *insulators*.

Positive and Negative Electricity

In the course of his experiments du Fay also discovered that there were two kinds of electrification: he showed that electrified glass and amber tended to oppose one another's attractiveness. To illustrate how he did so, we may use ebonite instead of amber, which has the same electrical properties. We suspend a pith-ball, and attract it with an electrified ebonite rod E (Fig. 30.1(i)); we then bring an electrified glass rod G towards the ebonite rod, and the pith-ball falls away (Fig. 30.1(ii)). Benjamin Franklin, a pioneer of electrostatics, gave the name of 'positive electricity' to the charge on a glass rod rubbed with silk, and 'negative electricity' to that on an ebonite rod rubbed with fur.

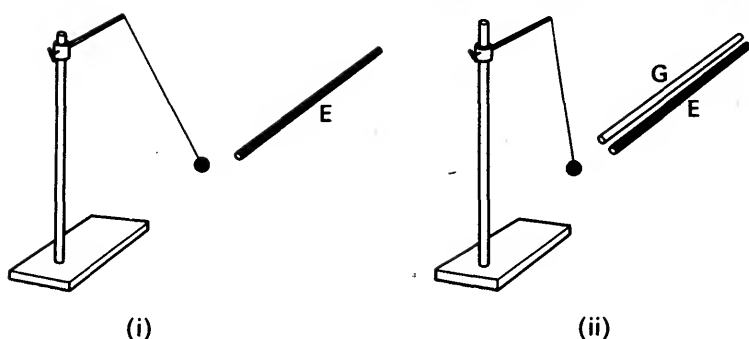


FIG. 30.1. Demonstrating that electrified glass tends to oppose effect of electrified ebonite.

Electrons and Electrostatics

Towards the end of the nineteenth century Sir J. J. Thomson discovered the existence of the *electron* (p. 1002). This is the lightest particle known—it is about $1/1840$ th of the mass of the hydrogen atom—and experiments show that it carries a tiny quantity of *negative* electricity. Later experiments showed that electrons are present in all atoms.

The detailed structure of atoms is complicated, but, generally, electrons exist round a minute core or nucleus carrying positive electricity. Normally, atoms are electrically neutral, that is, there is no surplus of charge on them. Consequently the total negative charge on the electrons is equal to the positive charge on the nucleus. In insulators, all the electrons appear to be firmly 'bound' to the nucleus under the attraction of the unlike charges. In metals, however, some of the electrons appear to be relatively 'free'. These electrons play an important part in electrical phenomena concerning metals.

The theory of electrons (negatively charged particles) gives simple explanations of electrification by friction, and of the attraction of uncharged bodies by charged ones. If the silk on which a glass rod has been rubbed is brought near to a charged and suspended ebonite rod it repels it; the silk must therefore have a negative charge. We know that the glass has a positive charge, and therefore we suppose that when the two were rubbed together electrons from the surface atoms were transferred from the glass to the silk. Likewise we suppose that when fur and ebonite are rubbed together, electrons go from the fur to the ebonite.

Attraction of Charged Body for Uncharged Bodies

To explain the attraction of a charged body for an uncharged one, we shall suppose that the uncharged body is a conductor—a metal. If it is brought near to a charged ebonite rod, say, then the negative charge on the rod repels the free electrons in the metal to its remote end (Fig. 30.2). A positive charge is thus left on the near end of the metal; this, being nearer than the negative charge on the far end, is attracted more strongly than the negative charge is repelled. On the

whole, therefore, the metal is attracted. If the uncharged body is not a conductor, the mechanism by which it is attracted is more complicated; we shall postpone its description to a later chapter.



FIG. 30.2. Attraction by charged body.

Electrostatics Today

The discovery of the electron has led, in the last twenty or thirty years, to a great increase in the practical importance of electrostatics. In devices such as radio valves and cathode-ray tubes, for example, electrons are moving under the influence of electrostatic forces. The problems of preventing sparks and the breakdown of insulators are essentially electrostatic. There are also difficulties in making measurements at very high voltages. These problems occur in high-voltage electrical engineering. Later, we shall also describe a modern electrostatic generator, of the type used to provide a million volts or more for X-ray work and nuclear bombardment. Such generators work on principles of electrostatics discovered over a hundred years ago.

Gold-leaf Electroscope

One of the earliest instruments used for testing positive and negative charges consisted of a metal rod A to which gold leaves L were

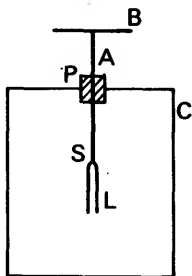


FIG. 30.3. A gold-leaf electroscope.

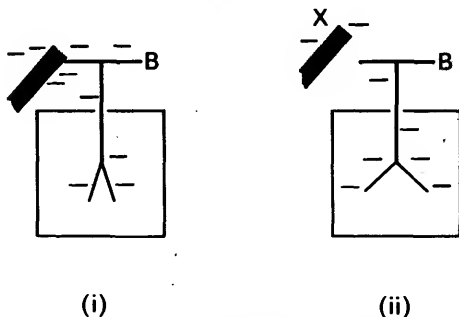


FIG. 30.4. Testing charge with electroscope.

attached (Fig. 30.3). The rod was fitted with a circular disc or cap B, and was insulated with a plug P from a metal case C which screened L from outside influences other than those brought near to B.

When B is touched by an ebonite rod rubbed with fur, some of the negative charge on the rod passes to the cap and L; and since like charges repel, the leaves diverge (Fig. 30.4(i)). If an unknown charge X is now brought near to B, an increased divergence implies that X is negative (Fig. 30.4(ii)). A positive charge is tested in a similar way; the electroscope is first given a positive charge and an increased divergence indicates a positive charge.

Induction

We shall now show that it is possible to obtain charges, called *induced charges*, without any contact with another charge. An experiment on electrostatic induction, as the phenomenon is called, is shown in Fig. 30.5(i). Two insulated metal spheres A, B are arranged so that they

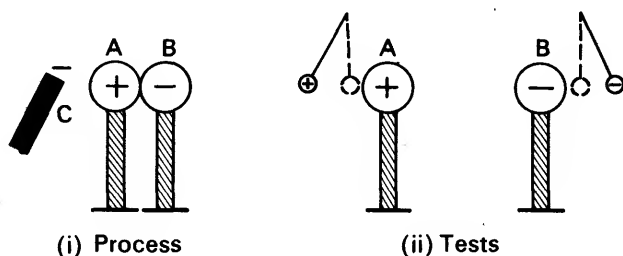


FIG. 30.5. Charges induced on a conductor.

touch one another, and a negatively charged ebonite rod C is brought near to A. The spheres are now separated, and then the rod is taken away. Tests with a charged pith-ball now show that A has a positive charge and B a negative charge (Fig. 30.5(ii)). If the spheres are placed together so that they touch, it is found that they now have no effect on a pith-ball held near. Their charges must therefore have neutralized each other completely, thus showing that the induced positive and negative charges are equal. This is explained by the movement of electrons from A to B when the rod is brought near. B has then a negative charge and A an equal positive charge.

Charging by Induction

Fig. 30.6 shows how a conductor can be given a permanent charge by induction, without dividing it in two. We first bring a charged ebonite rod, say, near to the conductor, (i); next we connect the conductor to

earth by touching it momentarily (ii); finally we remove the ebonite. We then find that the conductor is left with a positive charge (iii). If we use a charged glass rod, we find that the conductor is left with a negative charge; the charge left, called the induced charge, has always the opposite sign to the inducing charge.

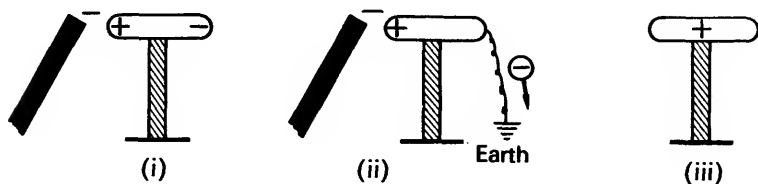


FIG. 30.6. Charging permanently by induction.

This phenomenon of induction can again be explained by the movement of electrons. If the inducing charge is negative, then, when we touch the conductor, electrons are repelled from it to earth, as shown in Fig. 30.6(ii), and a positive charge is left on the conductor. If the inducing charge is positive, then the electrons are attracted up from the earth to the conductor, which then becomes negatively charged.

Induction and the Electroscope

It is always observed that the leaves of an electroscope diverge when a charged body is brought near its cap, without touching it. This we can now easily understand; if, for example, we bring a negatively charged rod near the cap, it induces a positive charge on the cap, and a negative one on the leaves: the leaves then repel each other. Further, the negative charge on the leaves induces a positive one on the inside of the case, the corresponding negative charge running to the earth, on which the case rests. The positive charge on the case attracts the negative charge on the leaves, and makes them diverge further.

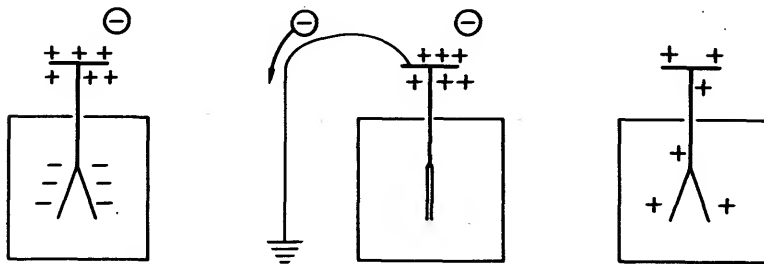


FIG. 30.7. Charging electroscope by induction.

We can use induction to give a permanent charge to the cap and leaves of an electroscope, by momentarily earthing the cap while holding an inducing charge near it. This is illustrated in Fig. 30.7.

The Electrophorus

A device which provides an almost unlimited supply of charge, by induction, was invented by Volta about 1800; it is called an *electrophorus*. It consists of an ebonite or perspex base, E in Fig. 30.8, and a metal disc D on an insulating handle. The ebonite is charged negatively by rubbing it—or, much better, beating it—with fur. The disc is then laid upon it, and acquires induced charges, positive underneath and negative on top, (i). Very little negative charge escapes from the ebonite to the disc, because the natural unevenness of their surfaces prevents them touching at more than a few points; charge escapes from these points only, because the ebonite is a non-conductor. After it has been placed on the ebonite, the disc is earthed with the finger, and the negative charge on its upper surface flows away, (ii). The disc can then be removed, and carries with it the positive charge which was on its underside, (iii).

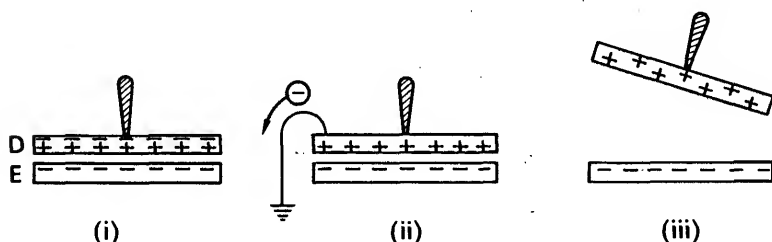


FIG. 30.8. The electrophorus.

An electrophorus produces sufficient charge to give an audible—and sometimes a visible—spark. The disc can be discharged and charged again repeatedly, until the charge on the ebonite has disappeared by leakage. Apparently, therefore, it is in principle an inexhaustible source of energy. However, work is done in raising the disc from the ebonite, against the attraction of their opposite charges, and this work must be done each time the disc is charged; the electrophorus is therefore not a source of energy, but a device for converting it from a mechanical into an electrical form.

The action of the electrophorus illustrates the advantages of charging by induction. First, the supply of charge is almost inexhaustible, because the original charge is not carried away. Second, a great charge—nearly equal to the charge on the whole of the ebonite—can be concentrated on to the conducting disc. As we have seen, only a very small charge could be transferred by contact, because the ebonite is not a conductor.

The Action of Points, Van de Graaff Generator

Sometimes in experiments with an electroscope connected to other apparatus by a wire, the leaves of the electroscope gradually collapse, as though its charge were leaking away. This behaviour can often be traced to a sharp point on the wire—if the point is blunted, the leakage stops. Charge leaks away from a sharp point through the air, being carried by molecules away from the point. This is explained later (p. 749).

Points are used to collect the charges produced in *electrostatic generators*. These are machines for induction, and thus building up very great charges and potential differences. Fig. 30.9 is a simplified diagram of one such machine, due to Van de Graaff. A hollow metal sphere S is supported on an insulating tube T , many feet high. A silk belt B runs over the pulleys shown, of which the lower is driven by an electric motor. Near the bottom and top of its run, the belt passes close to the electrodes E , which are sharply pointed combs, pointing towards the belt. The electrode E_1 is made about 10000 volts positive with respect to the earth by a battery. Its point then sprays the lower part of the belt with positive charge, which is carried up into the sphere. There it induces a negative charge on the points of electrode E_2 and a positive charge on the sphere to which the blunt end of E_2 is connected. The point sprays the belt with negative charge, and discharges it before it passes over the pulley. The sphere gradually charges up positively, until its potential is about a million volts relative to the earth.

Large machines of this type are used with high-voltage X-ray tubes, and for atom-splitting experiments. They have more elaborate electrode systems, stand about 15 m high, and have 4 m spheres. They can produce potential differences up to 5 000 000 volts and currents of about 50 microamperes. The electrical energy which they deliver comes from the work done by the motor in drawing the positively charged belt towards the positively charged sphere, which repels it.

In all types of high-voltage equipment sharp corners and edges must be avoided, except where points are deliberately used as electrodes. Otherwise, corona discharges may break out from the sharp places. All such places are therefore enlarged by metal globes, these are called stress-distributors. See also p. 749.

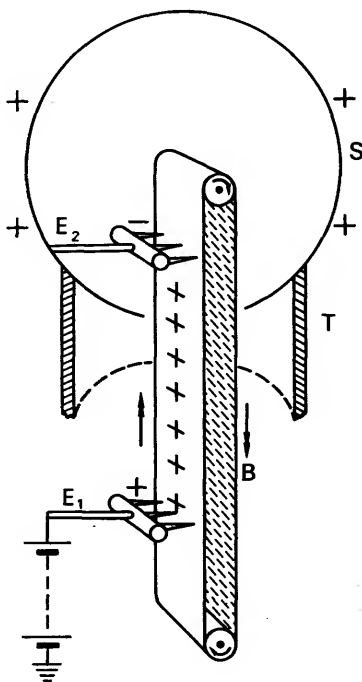


FIG. 30.9. Principle of Van de Graaff generator.



FIG. 30.10. Van de Graaff Electrostatic Generator at Aldermaston, England. The dome is the high-voltage terminal. The insulated rings are equipotentials, and provide a uniform potential gradient down the column. Beams of protons or deuterons, produced in the dome, are accelerated down the column to bombard different materials at the bottom, thereby producing nuclear reactions which can be studied.

Ice-pail Experiment

A famous experiment on electrostatic induction was made by Faraday in 1843. In it he used the ice-pail from which it takes its name; but it was a modest pail, 27 cm high—not a bucket. He stood the pail on an insulator, and connected it to a gold-leaf electroscope, as in Fig. 30.11(i). He next held a metal ball on the end of a long silk thread, and charged it positively by a spark from an electrophorus. Then he lowered the ball into the pail, without letting it touch the sides or bottom (Fig. 30.11(ii)).

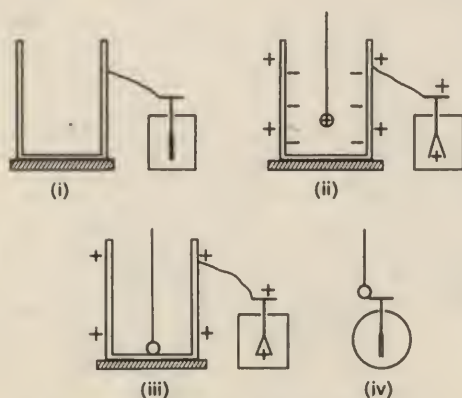


FIG. 30.11. Faraday's ice-pail experiment.

A positive charge was induced on the outside of the pail and the leaves, and made the leaves diverge. Once the ball was well inside the pail, Faraday found that the divergence of the leaves did not change when he moved the ball about—nearer to or farther from the walls or the bottom. This showed that the amount of the induced positive charge did not depend on the position of the ball, once it was well inside the pail. Faraday then allowed the ball to touch the pail, and noticed that the leaves of the electroscope still did not move (Fig. 30.11(iii)). When the ball touched the pail, therefore, no charge was given to, or taken from, the outside of the pail. Faraday next lifted the ball out of the pail, and tested it for charge with another electroscope. He found that the ball had no charge whatever (Fig. 30.11(iv)). The induced negative charge on the inside of the pail, must therefore have been equal in magnitude to the original positive charge on the ball.

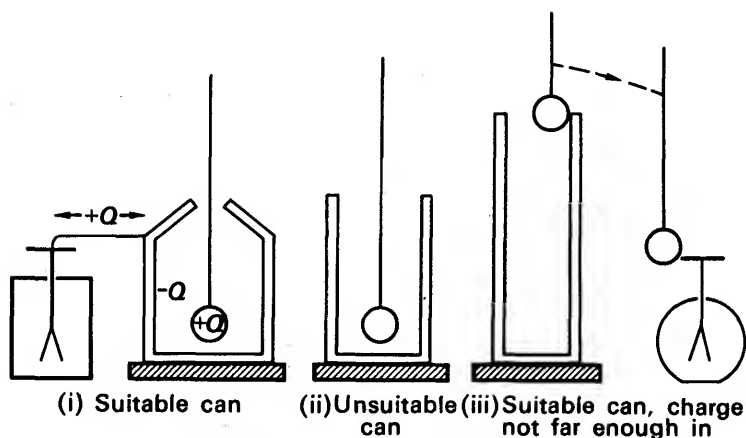


FIG. 30.12. Referring to Faraday's ice-pail experiment.

Faraday's experiment does not give these simple results unless the pail—or whatever is used in place of it—very nearly surrounds the charged ball (Fig. 30.12(i)(ii)). If, for example, the ball is allowed to touch the pail before it is well inside, as in Fig. 30.12(iii), then it does not lose all its charge.

Conclusions

The conclusions to be drawn from the experiment therefore apply, strictly, to a *hollow closed conductor*. They are:

(i) When a charged body is enclosed in a hollow conductor it induces on the inside of that conductor a charge equal but opposite to its own; and on the outside a charge equal and similar to its own (Fig. 30.11(i)).

(ii) The *total* charge inside a hollow conductor is always zero: either there are equal and opposite charges on the inside walls and within the volume (before the ball touches), or there is no charge at all (after the ball has touched).

Comparison and Collection of Charges

Faraday's ice-pail experiment gives us a method of comparing quantities of electric charges. The experiment shows that if a charged body is lowered well inside a tall, narrow can then it gives to the outside of the can a charge equal to its own. If the can is connected to the cap of an electroscope, the divergence of the leaves is a measure of the charge on the body. Thus we can compare the magnitudes of charges, without removing them from the bodies which carry them: we merely lower those bodies, in turn, into a tall insulated can, connected to an electroscope.

Sometimes we may wish to discharge a conductor completely, without letting its charge run to earth. We can do this by letting the conductor touch the bottom of a tall can on an insulating stand. The whole of the body's charge is then transferred to the outside of the can.

Charges Produced by Separation; Lines of Force

The ice-pail experiment suggests that a positive electric charge, for example, is always accompanied by an equal negative charge. Faraday repeated his experiment with a nest of hollow conductors, insulated

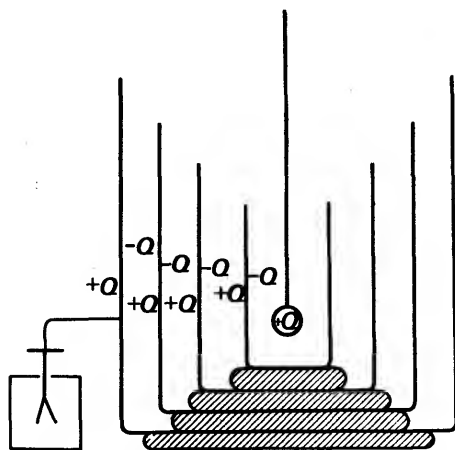


FIG. 30.13. Extension of the ice-pail experiment.

from one another, and showed that equal and opposite charges were induced on the inner and outer walls of each (Fig. 30.13).

Faraday also showed that equal and opposite charges are produced when a body is electrified by rubbing. He fitted an ebonite rod with a fur cap, which he rotated by a silk thread or string wrapped round it (Fig. 30.14(i)); he then compared the charges produced with an ice-pail and electroscope (Fig. 30.14(ii)(iii)(iv)).

In describing the conclusions from this last experiment, we now say, as indeed we have done already, that electrons flow from the fur to the

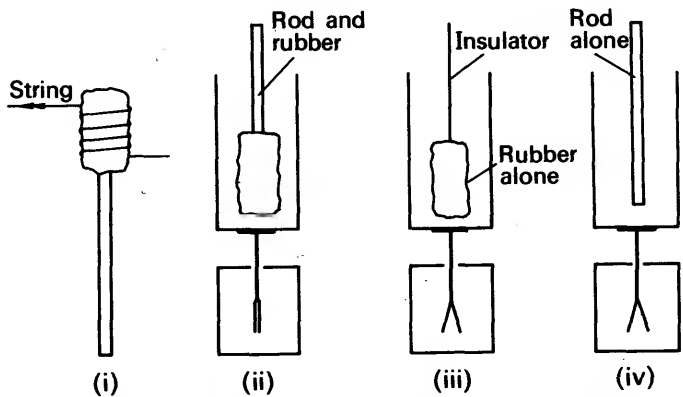


FIG. 30.14. Stages in showing that equal and opposite charges are produced by friction.

ebonite, carrying to it a negative charge, and leaving on the fur a positive charge. It appears, therefore, that free charges are always produced by separating equal amounts of the opposite kinds of electricity.

The idea that charges always occur in equal opposite pairs affects our drawing of lines of force diagrams. Lines of force radiate outwards

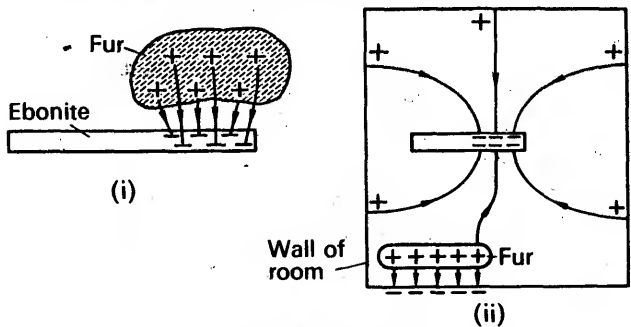


FIG. 30.15. Charging by friction—lines of force.

from a positive charge, and inwards to a negative one; from any positive charge, therefore, we draw lines of force ending on an equal negative charge. Figs. 30.15, 30.16 give some illustrations of this procedure.

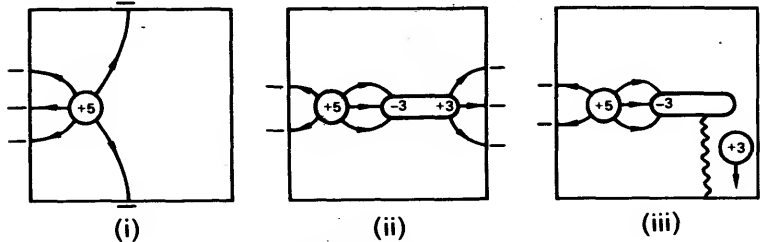


FIG. 30.16. Charging by induction—lines of force.

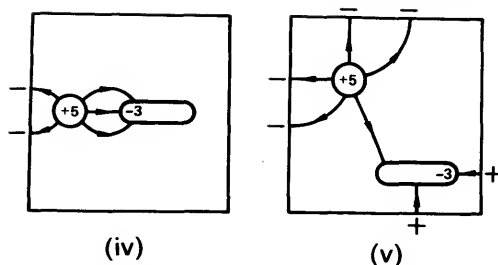


FIG. 30.16. Charging by induction—lines of force.

Distribution of Charge; Surface Density

By using a can connected to an electroscope we can find how electricity is distributed over a charged conductor of any form—pear-shaped, for example. We take a number of small leaves of tin-foil, all of the same area, but differently shaped to fit closely over the different parts of the conductor, and mounted on ebonite handles (Fig. 30.17(i)).

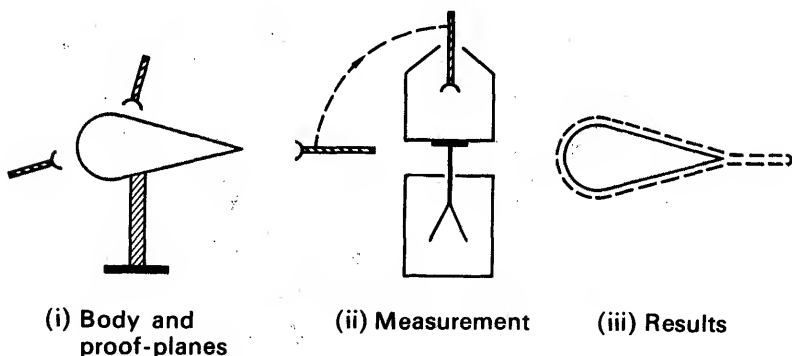


FIG. 30.17. Investigating charge distribution.

These are called *proof-planes*. We charge the body from an electrophorus, press a proof-plane against the part which it fits, and then lower the proof-plane into a can connected to an electroscope (Fig. 30.17(ii)). After noting the divergence of the leaves we discharge the can and electroscope by touching one of them, and repeat the observation with a proof-plane fitting a different part of the body. Since the proof-planes have equal areas, each of them carries away a charge proportional to the charge per unit area of the body, over the region which it touched. The charge per unit area over a region of the body is called the *surface-density* of the charge in that region. We find that the surface-density increases with the curvature of the body, as shown in Fig. 30.17(iii); the distance of the dotted line from the outline of the body is proportional to the surface-density of charge.

THE ELECTROSTATIC FIELD

Law of Force between two Charges

The magnitude of the force between two electrically charged bodies was studied by Coulomb in 1785. He showed that, if the bodies were small compared with the distance between them, then the force F was inversely proportional to the square of the distance r , i.e.

$$F \propto \frac{1}{r^2}. \quad (1)$$

This result is known as the *inverse square law*, or Coulomb's law.

It is not possible to verify the law accurately by direct measurement of the force between two charged bodies. In 1936 Plimton and Lawton showed, by an indirect method, that the power in the law cannot differ from 2 by more than $\pm 2 \times 10^{-9}$. We have no reason to suppose, therefore, that the inverse square law is other than exactly true.

Quantity of Charge

The SI unit of charge is the *coulomb* (C). The *ampere* (A), the unit of current, is defined later (p. 939). The coulomb is defined as that quantity of charge which passes a section of a conductor in one second when the current flowing is one ampère.

By measuring the force F between two charges when their respective magnitudes Q and Q' are varied, it is found that F is proportional to the product QQ' . Thus

$$F \propto QQ' \quad (2)$$

Law of Force

Combining (1) and (2), we have

$$F \propto \frac{QQ'}{r^2}$$

$$\therefore F = k \frac{QQ'}{r^2}, \quad (3)$$

where k is a constant. For reasons explained later, k is written as $1/4\pi\epsilon_0$, where ϵ_0 is a constant called the *permittivity* of free space if we suppose the charges are situated in a vacuum. Thus

$$F = \frac{1}{4\pi\epsilon_0} \frac{QQ'}{r^2} \quad (4)$$

In this expression, F is measured in newtons (N), Q in coulombs (C) and r in metres (m). Now, from (4),

$$\epsilon_0 = \frac{QQ'}{4\pi Fr^2}.$$

Hence the units of ϵ_0 are coulomb² newton⁻¹ metre⁻² (C² N⁻¹ m⁻²). Another unit of ϵ_0 , more widely used, is *farad metre*⁻¹ (F m⁻¹). See p. 774.

We shall see later that ϵ_0 has the numerical value of 8.854×10^{-12} , and $1/4\pi\epsilon_0$ then has the value 9×10^9 approximately.

Permittivity

So far we have considered charges in a vacuum. If charges are situated in other media such as water, then the force between the charges is reduced. Equation (4) is true only in a vacuum. In general, we write

$$F = \frac{1}{4\pi\epsilon} \frac{QQ'}{r^2} \quad (1)$$

where ϵ is the *permittivity* of the medium. The permittivity of air at normal pressure is only about 1.005 times that, ϵ_0 , of a vacuum. For most purposes, therefore, we may assume the value of ϵ_0 for the permittivity of air. The permittivity of water is about eighty times that of a vacuum. Thus the force between charges situated in water is eighty times less than if they were situated the same distance apart in a vacuum.

EXAMPLE

(a) Calculate the value of two equal charges if they repel one another with a force of 0.1 N when situated 50 cm apart in a vacuum.

(b) What would be the size of the charges if they were situated in an insulating liquid whose permittivity was ten times that of a vacuum?

(a) From (4),

$$F = \frac{1}{4\pi\epsilon_0} \frac{QQ'}{r^2}.$$

Since $Q = Q'$ here,

$$0.1 = \frac{9 \times 10^9 Q^2}{(0.5)^2}$$

or

$$Q^2 = \frac{0.1 \times (0.5)^2}{9 \times 10^9}$$

$$Q = 1.7 \times 10^{-6} \text{ C (coulomb), approx.}$$

$$= 1.7 \mu\text{C (microcoulomb).}$$

(b) The permittivity of the liquid $\epsilon = 10 \epsilon_0$.

$$F = \frac{1}{4\pi\epsilon} \frac{QQ'}{r^2}$$

$$= \frac{1}{10(4\pi\epsilon_0)} \frac{Q^2}{r^2}$$

$$Q^2 = \frac{(0.1) \times (0.5)^2 \times 10}{9 \times 10^9}$$

$$Q = 5.3 \times 10^{-6} \text{ C} = 5.3 \mu\text{C}.$$

Electric Intensity or Field-strength. Lines of Force

An 'electric field' can be defined as a region where an electric force is experienced. As in magnetism, electric fields can be mapped out by

electrostatic lines of force, which may be defined as a line such that the tangent to it is in the direction of the force on a small positive charge at that point. Arrows on the lines of force show the direction of the force on a positive charge; the force on a negative charge is in the opposite direction. Fig 30.18 shows the lines of force, also called *electric flux*, in some electrostatic fields.

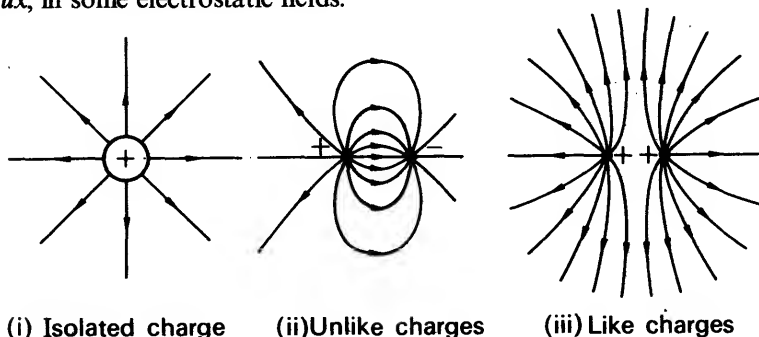


FIG. 30.18. Lines of electrostatic force.

The force exerted on a charged body in an electric field depends on the charge of the body and on the *intensity* or *strength* of the field. If we wish to explore the variation in intensity of an electric field, then we must place a test charge Q' at the point concerned which is small enough not to upset the field by its introduction. The intensity E of an electrostatic field at any point is defined as *the force per unit charge* which it exerts at that point. Its direction is that of the force exerted on a positive charge.

From this definition,

$$E = \frac{F}{Q'}$$

$$F = EQ' \quad (1)$$

Since F is measured in newtons and Q' in coulombs, it follows that *intensity E has units of newton per coulomb (N C^{-1})*. We shall see later that a more practical unit of E is *volt metre $^{-1}$ (V m^{-1})* (see p. 755).

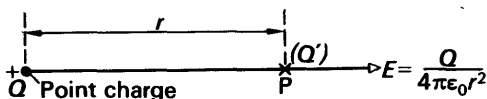


FIG. 30.19. Electric field intensity due to point charge.

We can easily find an expression for the strength E of the electric field due to a point charge Q situated in a vacuum (Fig. 30.19). We start from equation (4), p. 743, for the force between two such charges:

$$F = \frac{1}{4\pi\epsilon_0} \frac{QQ'}{r^2}$$

If the test charge Q' is situated at the point P in Fig. 30.19, the electric field strength at that point is given by (1).

$$\therefore E = \frac{F}{Q'} = \frac{Q}{4\pi\epsilon_0 r^2} \quad (2)$$

The direction of the field is radially outward if the charge Q is positive (Fig. 30.18(i)); it is radially inward if the charge Q is negative. If the charge were surrounded by a material of permittivity ϵ then,

$$E = \frac{Q}{4\pi\epsilon r^2} \quad (3)$$

Flux from a Point Charge

We have already shown how electric fields can be described by lines of force. From Fig. 30.18(i) it can be seen that the density of the lines increases near the charge where the field intensity is high. The intensity E at a point can thus be represented by the *number of lines per unit area* through a surface perpendicular to the lines of force at the point considered. The *flux* through an area perpendicular to the lines of force is the name given to the product of $E \times \text{area}$, where E is the intensity at that place. This is illustrated in Fig. 30.20(i).

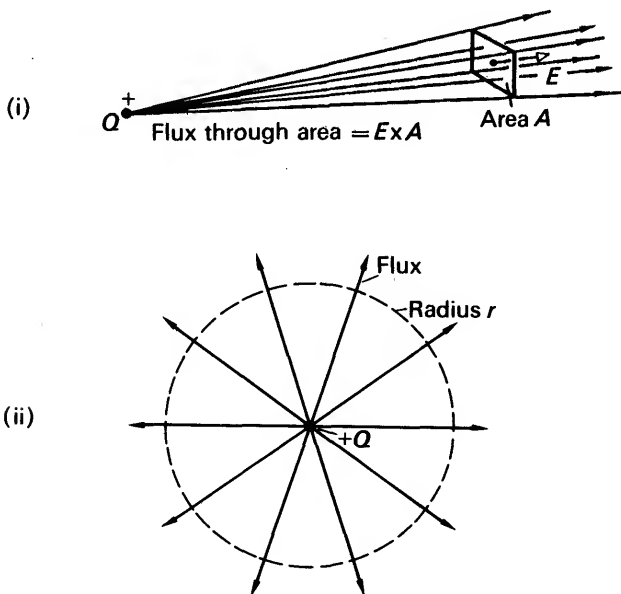


FIG. 30.20. Flux from a point charge.

Consider a sphere of radius r drawn in space concentric with a point charge (Fig. 30.20(ii)). The value of E at this place is given by (3),

p. 746. The total flux through the sphere is,

$$\begin{aligned}
 E \times \text{area} &= E \times 4\pi r^2 \\
 &= \frac{Q}{4\pi\epsilon r^2} \times 4\pi r^2 \\
 &= \frac{Q}{\epsilon} \\
 &= \frac{\text{charge inside sphere}}{\text{permittivity}} \quad (1)
 \end{aligned}$$

This demonstrates the important fact that the total flux crossing any sphere drawn outside and concentrically around a point charge is a constant. It does not depend on the distance from the charged sphere. It should be noted that this result is only true if the inverse square law is true.

To see this, suppose some other force law were valid, i.e. $E = Q/4\pi\epsilon r^n$. Then the total flux through the area

$$\begin{aligned}
 &= \frac{Q}{4\pi\epsilon r^n} \times 4\pi r^2 \\
 &= \frac{Q}{\epsilon} r^{(2-n)}
 \end{aligned}$$

This is only independent of r if $n = 2$.

Field due to Charged Sphere and Plane Conductor

Equation (1) can be shown to be generally true. Thus the flux passing through any *closed* surface whatever its shape, is always equal to Q/ϵ , where Q is the total charge enclosed by the surface. This relation, called *Gauss's Theorem*, can be used to find the value of E in other common cases.

(1) Outside a charged sphere

The flux across a spherical surface of radius r , concentric with a small sphere carrying a charge Q (Fig. 30.21), is given by,

$$\begin{aligned}
 \text{Flux} &= \frac{Q}{\epsilon} \\
 \therefore E \times 4\pi r^2 &= \frac{Q}{\epsilon} \\
 \therefore E &= \frac{Q}{4\pi\epsilon r^2}
 \end{aligned}$$

This is the same answer as that for a point charge. This means that *outside* a charged sphere, the field behaves as if all the charge on the sphere were concentrated at the centre.

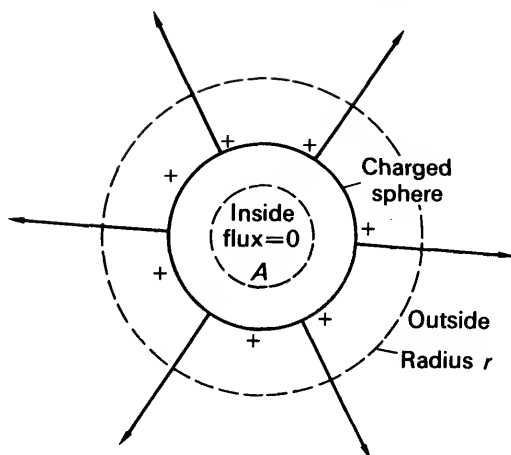


FIG. 30.21. Electric field of a charged sphere.

(2) *Inside a charged empty sphere*

Suppose a spherical surface A is drawn *inside* a charge sphere, as shown in Fig. 30.21. Inside this sphere there are no charges and so Q in equation (1), p. 747, is zero. This result is independent of the radius drawn, provided that it is less than that of the charged sphere. Hence from (1), p. 747, E must be zero everywhere inside a charged sphere.

(3) *Outside a Charged Plane Conductor*

Now consider a charged *plane* conductor S , with a surface charge density of σ coulomb metre⁻². Fig. 30.22 shows a plane surface P , drawn outside S , which is parallel to S and has an area A metre². Applying equation (1),

$$\therefore E \times \text{area} = \frac{\text{Charge inside surface}}{\epsilon}$$

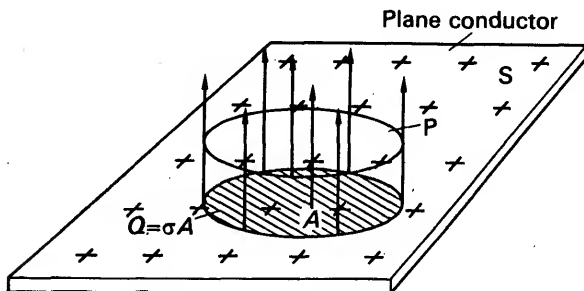


FIG. 30.22. Field of a charged plane conductor.

Now by symmetry, the intensity in the field must be perpendicular to the surface. Further, the charges which produce this field are those in the projection of the area P on the surface S , i.e. those within the shaded

area A in Fig. 30.22. The total charge here is thus σA coulomb.

$$\therefore E \cdot A = \frac{\sigma A}{\epsilon}$$

$$\therefore E = \frac{\sigma}{\epsilon}$$

Field Round Points

On p. 742 we saw that the surface-density of charge (charge per unit area) round a point of a conductor is very great. Consequently, the strength of the electric field near the point is very great. The intense electric field breaks down the insulation of the air, and sends a stream of charged molecules away from the point. The mechanism of the breakdown, which is called a 'corona discharge', is complicated, and we shall not discuss it here; some of the processes in it are similar to those in conduction through a gas at low pressure, which we shall describe in Chapter 40. Corona breakdown starts when the electric field strength is about 3 million volt metre⁻¹. The corresponding surface-density is about 2.7×10^{-5} coulomb metre⁻².

EXAMPLE

An electron of charge 1.6×10^{-19} C is situated in a uniform electric field of intensity 1200 volt cm⁻¹. Find the force on it, its acceleration, and the time it takes to travel 2 cm from rest (electronic mass, $m_e = 9.10 \times 10^{-31}$ kg).

Force on electron $F = eE$.

Now $E = 1200$ volt cm⁻¹ = 120 000 volt m⁻¹.

$$\begin{aligned}\therefore F &= 1.6 \times 10^{-19} \times 1.2 \times 10^5 \\ &= 1.92 \times 10^{-14} \text{ N (newton).}\end{aligned}$$

$$\begin{aligned}\text{Acceleration, } a &= \frac{F}{m} = \frac{1.92 \times 10^{-14}}{9.1 \times 10^{-31}} \\ &= 2.12 \times 10^{16} \text{ m s}^{-2} \text{ (metre second}^{-2}\text{)}\end{aligned}$$

Time for 2 cm travel is given by

$$\begin{aligned}s &= \frac{1}{2}at^2 \\ \therefore t &= \sqrt{\frac{2s}{a}} = \sqrt{\frac{2 \times 0.02}{2.12 \times 10^{16}}} \\ &= 1.37 \times 10^{-9} \text{ seconds.}\end{aligned}$$

The extreme shortness of this time is due to the fact that the ratio of charge-to-mass for an electron is very great:

$$\frac{e}{m} = \frac{1.6 \times 10^{-19}}{9.1 \times 10^{-31}} = 1.8 \times 10^{11} \text{ C kg}^{-1}.$$

In an electric field, the charge e determines the force on an electron, while the mass m determines its inertia. Because of the large ratio e/m , the electron moves almost instantaneously, and requires very little energy to displace it. Also it can respond to changes in an electric field which take place even millions of times per second. Thus it is the large value of e/m for electrons which makes electronic tubes, for example, useful in electrical communication and remote control.

ELECTRIC POTENTIAL

Potential in Fields

When an object is held at a height above the earth it is said to have potential energy. A heavy body tends to move under the force of attraction of the earth from a point of great height to one of less, and we say that points in the earth's gravitational field have potential values depending on their height.

Electric potential is analogous to gravitational potential, but this time we think of points in an electric field. Thus in the field round a positive charge, for example, a positive charge moves from points near the charge to points further away. Points round the charge are said to have an 'electric potential'.

Potential Difference

In mechanics we are always concerned with differences of height; if a point A on a hill is h metre higher than a point B , and our weight is w newton, then we do wh joule of work in climbing from B to A (Fig. 30.23 (i)). Similarly in electricity we are often concerned with differences of potential; and we define these also in terms of work.

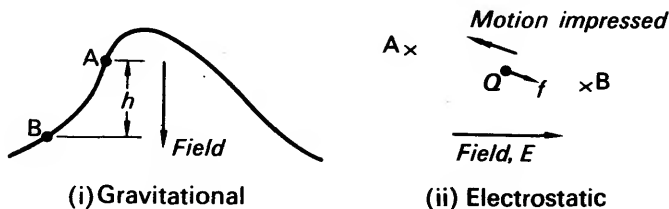


FIG. 30.23. Work done, in gravitational and electrostatic fields.

Let us consider two points A and B in an electrostatic field, and let us suppose that the force on a positive charge Q has a component f in the direction AB (Fig. 30.23 (ii)). Then if we move a positively charged body from B to A , we do work against this component of the field E . We define the potential difference between A and B as the work done in moving a unit positive charge from B to A . We denote it by the symbol V_{AB} .

The work done will be measured in joules (J). The unit of potential difference is called the volt and may be defined as follows: The potential difference between two points A and B is one volt if the work done in taking one coulomb of positive charge from B to A is one joule.

From this definition, if a charge of Q coulombs is moved through a p.d. of V volt, then the work done W in joules is given by

$$W = QV \quad (1)$$

Potential and Energy

Let us consider two points A and B in an electrostatic field, A being at a higher potential than B . The potential difference between A and

B we denote as usual by V_{AB} . If we take a positive charge Q from B to A, we do work on it of amount QV_{AB} : the charge gains this amount of potential energy. If we now let the charge go back from A to B, it loses that potential energy: work is done on it by the electrostatic force, in the same way as work is done on a falling stone by gravity. This work may become kinetic energy, if the charge moves freely, or external work if the charge is attached to some machine, or a mixture of the two.

The work which we must do in first taking the charge from B to A does not depend on the path along which we carry it, just as the work done in climbing a hill does not depend on the route we take. If this were not true, we could devise a perpetual motion machine, in which we did less work in carrying a charge from B to A via X than it did for us in returning from A to B via Y (Fig. 30.24).

The fact that the potential differences between two points is a constant, independent of the path chosen between the points, is the most important property of potential in general; we shall see why later on. This property can be conveniently expressed by saying that the work done in carrying a charge round a closed path in an electrostatic field, such as BXAYB in Fig. 30.24 is zero.

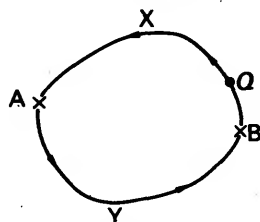


FIG. 30.24. A closed path in an electrostatic field.

Potential Difference Formula

To obtain a formula for potential difference, let us calculate the potential difference between two points in the field of a single point positive charge, Q in Fig. 30.25. For simplicity we will assume that the points, A and B, lie on a line of force at distances a and b respectively

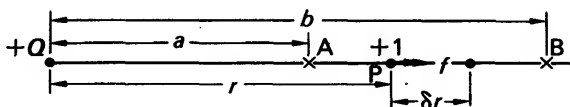


FIG. 30.25. Calculation of potential.

from the charge. When a unit positive charge is at a distance r from the charge Q in free space the force on it is

$$f = \frac{Q \times 1}{4\pi\epsilon_0 r^2}$$

The work done in taking the charge from B to A, against the force f , is equal to the work which the force f would do if the charge were allowed to go from A to B. Over the short distance δr , the work done by the force f is

$$\delta W = f\delta r.$$

Over the whole distance AB, therefore, the work done by the force on the unit charge is

$$\begin{aligned}\int_A^B \delta W &= \int_{r=a}^{r=b} f dr = \int_a^b \frac{Q}{4\pi\epsilon_0 r^2} dr \\ &= -\left[\frac{Q}{4\pi\epsilon_0 r} \right]_a^b = \frac{Q}{4\pi\epsilon_0 a} - \frac{Q}{4\pi\epsilon_0 b}\end{aligned}$$

This, then, is the value of the work which an external agent must do to carry a unit positive charge from B to A. The work per coulomb is the potential difference V_{AB} between A and B.

$$\therefore V_{AB} = \frac{Q}{4\pi\epsilon_0} \left(\frac{1}{a} - \frac{1}{b} \right) \quad (1)$$

V_{AB} will be in volts if Q is in coulombs, a and b are in metres and ϵ_0 is taken as 8.85×10^{-12} or $1/4\pi\epsilon_0$ as 9×10^9 approximately (see p. 744).

EXAMPLE

Two positive point charges, of 12 and 8 microcoulomb respectively, are 10 cm apart. Find the work done in bringing them 4 cm closer. (Assume $1/4\pi\epsilon_0 = 9 \times 10^9$.)

Suppose the 12 μC (microcoulomb) charge is fixed in position. Since 6 cm = 0.06 m and 10 cm = 0.1 m, then the potential difference between points 6 and 10 cm from it is given by (1).

$$\begin{aligned}\therefore V &= \frac{12 \times 10^{-6}}{4\pi\epsilon_0} \left(\frac{1}{0.06} - \frac{1}{0.1} \right) \\ &= 12 \times 10^{-6} \times 9 \times 10^9 (16\frac{2}{3} - 10) \\ &= 720\,000 \text{ V.}\end{aligned}$$

(Note the very high potential difference due to quite small charges.)

The work done in moving the 8 μC charge from 10 cm to 6 cm away from the other is given by, using $W = QV$,

$$\begin{aligned}W &= 8 \times 10^{-6} \times V \\ &= 8 \times 10^{-6} \times 720\,000 \\ &= 5.8 \text{ J.}\end{aligned}$$

Zero Potential

Instead of speaking continually of potential differences between pairs of points, we may speak of the potential at a single point—provided we always refer it to some other, agreed, reference point. This procedure is analogous to referring the heights of mountains to sea-level.

For practical purposes we generally choose as our reference point the electric potential of the surface of the earth. Although the earth is large it is all at the same potential, because it is a good conductor of electricity; if one point on it were at a higher potential than another, electrons would flow from the lower to the higher potential. As a result,

the higher potential would fall, and the lower would rise; the flow of electricity would cease only when the potentials became equalized.

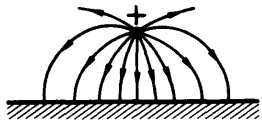


FIG. 30.26. Electric field of positive charge near earth.

In general it is difficult to calculate the potential of a point relative to the earth. This is because the electric field due to a charged body near a conducting surface is complicated, as shown by the lines of force diagram in Fig. 30.26. In theoretical calculations, therefore, we often find it convenient to consider charges so far from the earth that the effect of the earth on their field is negligible; we call these 'isolated' charges.

Thus we define the potential at a point A as V volts if V joules of work is done in bringing one coulomb of positive charge from infinity to A.

Potential Formula

Equation (1), p. 752, gives the potential difference between two points in the field of an isolated point charge Q . If we let the point B retreat to infinity, then $b \gg a$, and the equation gives for the potential at A:

$$V_A = \frac{Q}{4\pi\epsilon_0 a} \quad (1)$$

The derivation of this equation shows us what we mean by the word 'infinity': the distance b is infinite if $1/b$ is negligible compared with $1/a$. If a is 1 cm, and b is 1 m, we make an error of only 1 per cent. in ignoring it; if b is 100 m, then for all practical purposes the point B is at infinity. In atomic physics, where the distances concerned have the order of 10^{-8} cm, a fraction of a millimetre is infinite.

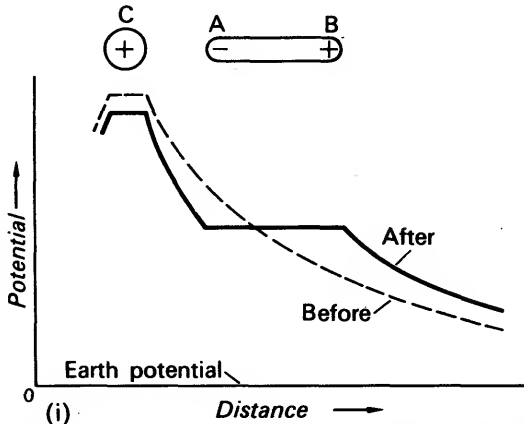


FIG. 30.27 (i). Potential distribution near a positive charge before and after bringing up an uncharged conductor.

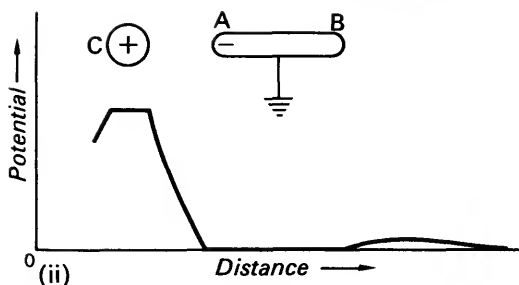


FIG. 30.27 (ii). Potential distribution near a positive charge in the presence of an earthed conductor.

In the neighbourhood of an isolated negative charge, the potential is negative, because Q in equation (1) is negative. The potential is also negative in the neighbourhood of a negative charge near the earth: the earth is at zero potential, and a positive charge will tend to move from it towards the negative charge. A negative potential is analogous to the depth of a mine below sea-level. Fig. 30.27 (i) shows the potential variation near a positive charge C before and after a conductor AB is brought near. Fig. 30.27 (ii) shows the potential variation when AB is earthed.

Potential Difference and Intensity

We shall now see how potential difference is related to intensity or field-strength. Suppose A , B are two neighbouring points on a line of force, so close together that the electric field-intensity between them is constant and equal to E (Fig. 30.28). If V is the potential at A , $V + \delta V$ is that at B , and the respective distances of A , B from the origin are x and $x + \delta x$, then

$$\begin{aligned} V_{AB} &= \text{potential difference between } A, B \\ &= V_A - V_B = V - (V + \delta V) = -\delta V. \end{aligned}$$

The work done in taking a unit charge from B to A

$$= \text{force} \times \text{distance} = E \times \delta x = V_{AB} = -\delta V.$$

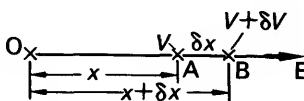


FIG. 30.28. Field strength and potential gradient.

Hence

$$E = -\frac{\delta V}{\delta x},$$

or, in the limit,

$$E = -\frac{dV}{dx}. \quad (1)$$

The quantity dV/dx is the rate at which the potential rises with distance, and is called the potential gradient. Equation (1) shows that the strength of the electric field is equal to the negative of the potential gradient, and strong and weak fields in relation to potential are illustrated in Fig. 30.29.

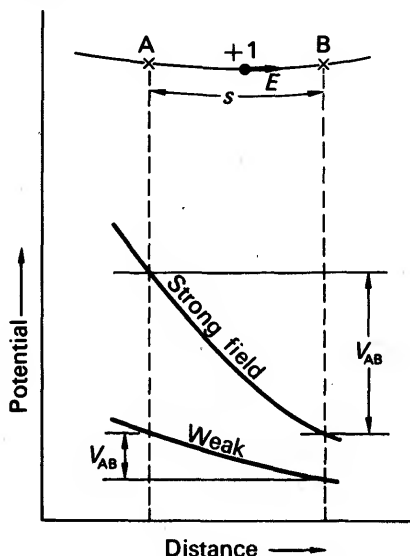


FIG. 30.29. Relationship between potential and field strength.

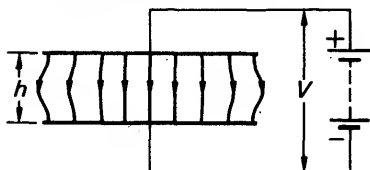


FIG. 30.30. Electric field between parallel plates.

In Fig. 30.30 the electric intensity $= V/h$, the potential gradient, and this is uniform in magnitude in the middle of the plates. At the edge of the plates the field becomes non-uniform.

We can now see why E is usually given in units of 'volt per metre' (V m^{-1}).

From (1), $E = -(dV/dx)$. Since V is measured in volts and x in metres, then E will be in volt per metre (V m^{-1}). From the original definition of E , summarized by equation (1) on p. 745, the units of E were newton coulomb $^{-1}$. To show that these are equivalent, from (1),

$$\begin{aligned} 1 \text{ volt} &= 1 \text{ joule coulomb}^{-1} \\ &= 1 \text{ newton metre coulomb}^{-1} \end{aligned}$$

Since

$$1 \text{ joule} = 1 \text{ newton} \times 1 \text{ metre}$$

$$\therefore 1 \text{ volt metre}^{-1} = 1 \text{ newton coulomb}^{-1}$$

EXAMPLES

1 An electron is liberated from the lower of two large parallel metal plates separated by a distance $h = 2$ cm. The upper plate has a potential of 2400 volts relative to the lower. How long does the electron take to reach it?

Between large parallel plates, close together, the electric field is uniform except near the edges of the plates, as shown in Fig. 30.30. Except near the edges, therefore, the potential gradient between the plates is uniform; its magnitude is V/h ,

$$\begin{aligned}\text{electric intensity } E &= \text{potential gradient} \\ &= 2400/0.02 \text{ V m}^{-1} \\ &= 1.2 \times 10^5 \text{ V m}^{-1}.\end{aligned}$$

The rest of this problem may now be worked out exactly as the example on p. 749.

2 An electron is liberated from a hot filament, and attracted by an anode, of potential 1200 volts positive with respect to the filament. What is the speed of the electron when it strikes the anode?

$$e = \text{electronic charge} = 1.6 \times 10^{-19} \text{ C.}$$

$$V = 1200 \text{ V, } m = \text{mass of electron} = 9.1 \times 10^{-31} \text{ kg.}$$

The energy which the electron gains from the field $= QV = eV$.

Kinetic energy gained

$$= \frac{1}{2}mv^2 = eV,$$

where v is the speed gained from rest.

$$\begin{aligned}v &= \sqrt{\frac{2eV}{m}} = \sqrt{\frac{2 \times 1.6 \times 10^{-19} \times 1200}{9.1 \times 10^{-31}}} \\ &= 2.1 \times 10^7 \text{ m s}^{-1}.\end{aligned}$$

The Electron-Volt

The kinetic energy gained by an electron which has been accelerated through a potential difference of 1 volt is called an *electron-volt* (eV). Since the energy gained in moving a charge Q through a p.d. $V = QV$,

$$\therefore 1 \text{ eV} = \text{electronic charge} \times 1 = 1.6 \times 10^{-19} \times 1 \text{ joule} = 1.6 \times 10^{-19} \text{ J.}$$

The electron-volt is a useful unit of energy in atomic physics. For example, the work necessary to extract a conduction electron from tungsten is 4.52 electron-volt. This quantity determines the magnitude of the thermionic emission from the metal at a given temperature (p. 1026); it is analogous to the latent heat of evaporation of a liquid.

EQUIPOTENTIALS

Equipotentials

We have already said that the earth must have the same potential all over, because it is a conductor. In a conductor there can be no differences of potential, because these would set up a potential gradient or electric field; electrons would then redistribute themselves throughout the conductor, under the influence of the field, until they had

destroyed the field. This is true whether the conductor has a net charge, positive or negative, or whether it is uncharged; it is true whatever the actual potential of the conductor, relative to any other body.

Any surface or volume over which the potential is constant is called an *equipotential*. The volume or surface may be that of a material body, or simply a surface or volume in space. For example, as we shall see later, the space inside a hollow charged conductor is an equipotential volume. Equipotential surfaces can be drawn throughout any space in which there is an electric field, as we shall now explain.

Let us consider the field of an isolated point-charge Q . At a distance a from the charge, the potential is $Q/4\pi\epsilon_0 a$; a sphere of radius a and centre at Q is therefore an equipotential surface, of potential $Q/4\pi\epsilon_0 a$. In fact, all spheres centred on the charge are equipotential surfaces, whose potentials are inversely proportional to their radii (Fig. 30.31). An equipotential surface has the property that, along any direction lying in the surface, there is no electric field; for there is no potential gradient. *Equipotential surfaces are therefore always at right angles to lines of force*, as shown in Fig. 30.31. This also shows numerical values proportional to their potentials. Since conductors are always equipotentials, if any conductors appear in an electric-field diagram the lines of force must always be drawn to meet them at right angles.

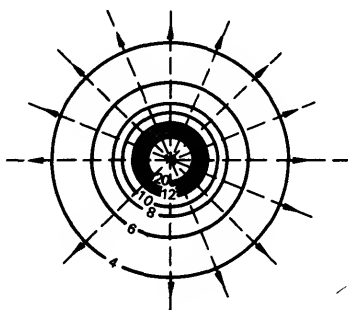


FIG. 30.31. Equipotentials and lines of force around a point charge.

Potential due to a System of Charges

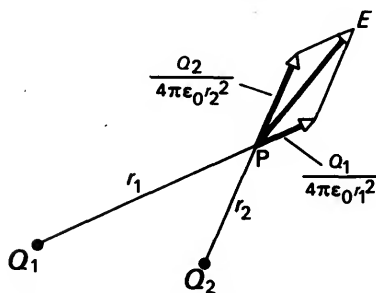
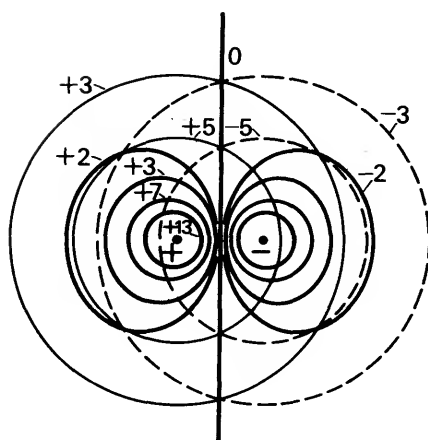


FIG. 30.32. Finding resultant field of two point-charges.

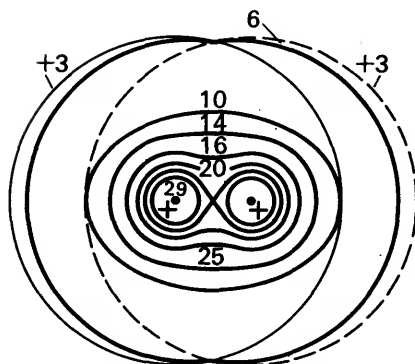
When we set out to consider the electric field due to more charges than one, then we see the advantages of the idea of potential over the idea of field-strength. If we wish to find the field-strength E at the point P in Fig. 30.32, due to the two charges Q_1 and Q_2 , we have first to find the force exerted by each on a unit charge at P , and then to compound these forces by the parallelogram method. See Fig. 30.32. On the other hand, if we wish to find the

potential at P , we merely calculate the potential due to each charge, and *add the potentials algebraically*.

Quantities which can be added algebraically are called 'scalars'; they may have signs—positive or negative, like a bank balance—but



(i) Opposite charges



(ii) Like charges

FIG. 30.33. Equipotentials in the field of two point charges.

they have no direction: they do not point north, east, south, or west. Quantities which have direction, like forces, are called 'vectors'; they have to be added either by resolution into components, or by the parallelogram method. Either way is slow and clumsy, compared with the addition of scalars. For example, we can draw the equipotentials round a point-charge with compasses; if we draw two sets of them, as in Fig. 30.33(i) or (ii), then by simple addition we can rapidly sketch the equipotentials around the two charges together.

And when we have plotted the equipotentials, they turn out to be more useful than lines of force. A line of force diagram appeals to the imagination, and helps us to see what would happen to a charge in the field. But it tells us little about the strength of the field—at the best, if it is more carefully drawn than most, we can only say that the field is strongest where the lines are closest. But equipotentials can

be labelled with the values of potential they represent ; and from their spacing we can find the actual value of the potential gradient, and hence the field-strength. The only difficulty in interpreting equipotential diagrams lies in visualizing the direction of the force on a charge ; this is always at right angles to the curves.

Field inside Hollow Conductor. Potential Difference and Gold-leaf Electroscope

If a hollow conductor contains no charged bodies, then, whatever charge there may be on its outside, there is none on its inside. Inside it, therefore, there is no electric field ; the space within the conductor is an equipotential volume. If the conductor has an open end, like a can, then most of the space inside it is equipotential, but near its mouth there is a weak field (Fig. 30.34).

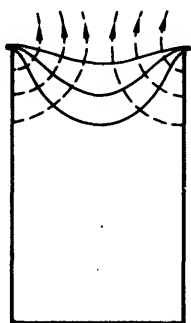


FIG. 30.34. Equipotentials and lines of force near mouth of an open charged can.

The behaviour of the *gold-leaf electroscope* illustrates this point. If we stand the case on an insulator, and connect the cap to it with a wire, then, no matter what charge we give to the cap, the leaves do not diverge (Fig. 30.35). Any charge we give to the cap spreads over the case

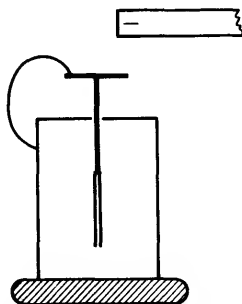


FIG. 30.35. Electroscope cap joined to case.

of the electroscope, but none appears on the leaves, and there is no force acting to diverge them. When, as usual, the cap is insulated and

the case earthed, charging the cap sets up a potential difference between it and the case. Charges appear on the leaves, and the field between them and the case makes them diverge (p. 733). If the case is insulated from earth, as well as from the cap, the leaves diverge less; the charge on them and the cap raises the potential of the case and reduces the potential difference between it and the leaves. The field acting on the leaves is thus made weaker, and the force on the leaves less. We can sum up these observations by saying that *the electroscope indicates the potential difference between its leaves and its case.*

Potential of Pear-shaped Conductor

On p. 742 we saw that the surface-density of the charge on a pear-shaped conductor was greatest where the curvature was greatest. The potential of the conductor at various points can be examined by means of the gold-leaf electroscope, the case being earthed. One end of a wire is connected to the cap; some of the wire is then wrapped round an insulating rod, and the free end of the wire is placed on the conductor. As the free end is moved over the conductor, it is observed that the divergence of the leaf remains constant. This result was explained on pp. 756, 757.

Electrostatic Shielding

The fact that there is no electric field inside a close conductor, when it contains no charged bodies, was demonstrated by Faraday in a spectacular manner. He made for himself a large wire cage, supported it on insulators, and sat inside it with his electroscopes. He then had the cage charged by an induction machine—a forerunner of the type we described on p. 737—until painful sparks could be drawn from its outside. Inside the cage Faraday sat in safety and comfort, and there was no deflection to be seen on even his most sensitive electroscope.

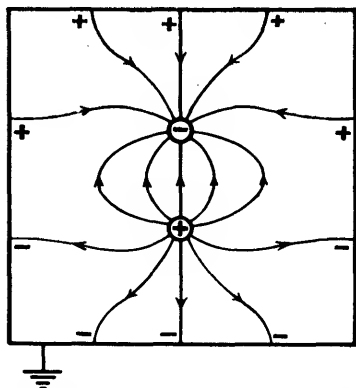


FIG. 30.36. Lines of force round charges.

If we wish to protect any persons or instruments from intense electric fields, therefore, we enclose them in hollow conductors; these are called 'Faraday cages', and are widely used in high-voltage measurements in industry.

We may also wish to prevent charges in one place from setting up an electric field beyond their immediate neighbourhood. To do this we surround the charges with a Faraday cage, and connect the cage to earth (Fig. 30.36). The charge induced on the outside of the cage then runs to earth, and there is no external field. (When a cage is used to shield something *inside* it, it does not have to be earthed.)

Comparison of Static and Current Phenomena

Broadly speaking, we may say that in electrostatic phenomena we meet small quantities of charge, but great differences of potential. On the other hand in the phenomena of current electricity discussed later, the potential differences are small but the amounts of charge transported by the currents are great. Sparks and shocks are common in electrostatics, because they require great potential differences; but they are rarely dangerous, because the total amount of energy available is usually small. On the other hand, shocks and sparks in current electricity are rare, but, when the potential difference is great enough to cause them, they are likely to be dangerous.

These quantitative differences make problems of insulation much more difficult in electrostatic apparatus than in apparatus for use with currents. The high potentials met in electrostatics make leakage currents relatively great, and the small charges therefore tend to disappear rapidly. Any wood, for example, ranks as an insulator for current electricity, but a conductor in electrostatics. In electrostatic experiments we sometimes wish to connect a charged body to earth; all we have then to do is to touch it.

EXAMPLE

Three charges $-1 \mu\text{C}$, $2 \mu\text{C}$ and $3 \mu\text{C}$ are placed respectively at the corners A, B, C of an equilateral triangle of side 2 metres. Calculate (a) the potential, (b) the electric field, at a point X which is half-way along BC. Fig. 30.37.

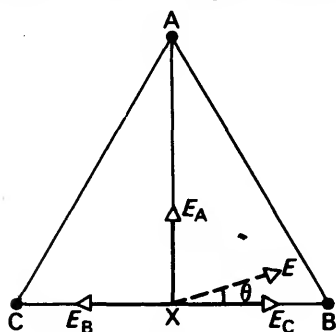


FIG. 30.37. Example.

(a) Potential at X due to charge at B is

$$\begin{aligned} \frac{Q}{4\pi\epsilon_0 r} &= \frac{2 \times 10^{-6}}{4\pi\epsilon_0 \times 1} \\ &= 18 \times 10^3 \text{ V.} \end{aligned}$$

Similarly potential at X due to the charge at C = $27 \times 10^3 \text{ V}$, and the potential due to A is

$$\begin{aligned} V_A &= \frac{Q}{4\pi\epsilon_0 r} = \frac{-10^{-6}}{4\pi\epsilon_0 \times \sqrt{3}} \\ &= -5 \times 10^3 \text{ V (approx.).} \end{aligned}$$

Since potential is a scalar quantity and can be added algebraically, the net potential at X

$$\begin{aligned} &= (18 + 27 - 5) \times 10^3 \text{ V} \\ &= 40 \times 10^3 \text{ V.} \end{aligned}$$

(b) The resultant field at X is due to the three electric fields from the three charges.

The field due to B, E_B has magnitude given by

$$\begin{aligned} E_B &= \frac{Q}{4\pi\epsilon_0 r^2} = \frac{2 \times 10^{-6}}{4\pi\epsilon_0 \times 1^2} \\ &= 18 \times 10^3 \text{ V m}^{-1}. \end{aligned}$$

Similarly

$$E_C = 27 \times 10^3 \text{ V m}^{-1}.$$

Since these act along the same straight line the resultant of E_B and $E_C = 9 \times 10^3 \text{ V m}^{-1}$ directed from C to B.

$$\begin{aligned}\text{Also, } E_A &= \frac{Q}{4\pi\epsilon_0 r^2} = \frac{10^{-6}}{4\pi\epsilon_0 \times (\sqrt{3})^2} \\ &= 3 \times 10^3 \text{ V m}^{-1}.\end{aligned}$$

The resultant field has magnitude, E , given by

$$\begin{aligned}E^2 &= E_A^2 + (E_C - E_B)^2 \\ &= (9 + 81) \times 10^6 \\ &= 90 \times 10^6 \\ E &= 3\sqrt{10} \times 10^3 \text{ V m}^{-1} = 9.5 \times 10^3 \text{ V m}^{-1}.\end{aligned}$$

This makes an angle θ with CB, in the direction shown by the dotted line, where

$$\tan \theta = \frac{E_A}{E_C - E_B} = \frac{3 \times 10^3}{9 \times 10^3} = \frac{1}{3}$$

$$\therefore \theta = 18^\circ 25'.$$

EXERCISES 30

- Describe experiments with a gold leaf electroscope:

- to demonstrate that this instrument indicates the potential difference between its leaves and its case and not necessarily the total charge on its leaves and cap;
- to compare the quantities of electricity on two conductors of unequal size;
- to investigate the distribution of electricity on a charged conductor.

In case (c) state the results you would expect if the conductor were spherical with a pointed rod attached to it, and describe and explain two practical applications of pointed conductors. (L.)

- Describe, with the aid of a labelled diagram, a Van de Graaff generator, explaining the physical principles of its action.

The high voltage terminal of such a generator consists of a spherical conducting shell of radius 50 cm. Estimate the maximum potential to which it can be raised in air for which electrical breakdown occurs when the electric intensity exceeds $30000 \text{ volt cm}^{-1}$.

State two ways in which this maximum potential could be increased. (N.)

- Define *potential at a point* in an electric field.

Sketch a graph illustrating the variation of potential along a radius from the centre of a charged isolated conducting sphere to infinity.

Assuming the expression for the potential of a charged isolated conducting sphere in air, determine the change in the potential of such a sphere caused by surrounding it with an earthed concentric thin conducting sphere having three times its radius. (N.)

- What is meant by the terms (a) *potential* and (b) *field strength* in electrostatics? State whether each quantity is a scalar or a vector.

Write down the law which gives the force between two point charges q_1 and q_2 at a distance r apart and use it to derive the electric field strength and the potential due to a point charge q at a distance x from it.

The points A, B and C form an equilateral triangle of side z . Point charges

of equal magnitude q are placed at A and B. Find the electric field strength and the potential at C due to these charges when (i) both charges are positive and (ii) the charge at A is positive and the charge at B is negative.

O is the midpoint of AB and POQ is the perpendicular bisector of AB; PO and OQ are very large distances compared with AB. Draw rough graphs to show how the magnitude of the electric field strength and the potential vary along the line POQ in cases (i) and (ii) above. (O. & C.)

5. Describe simple electrostatic experiments to illustrate two of the following: (a) the production of equal and opposite charges by induction; (b) the action of points; (c) the effect on a charged conductor of the approach of an earthed conductor. (L.)

6. Discuss the following, giving examples: *Electrostatic induction*, *electric discharge from points* and *dielectric strength*.

Describe a modern form of apparatus for obtaining a small current at a very high voltage. (L.)

7. An isolated conducting spherical shell of radius 10 cm, in vacuo, carries a positive charge of 1.0×10^{-7} coulomb. Calculate (a) the electric field intensity, (b) the potential, at a point on the surface of the conductor. Sketch a graph to show how one of these quantities varies with distance along a radius from the centre to a point well outside the spherical shell. Point out the main features of the graph. (N.)

8. Define (a) electric intensity, (b) difference of potential. How are these quantities related?

A charged oil-drop of radius 0.00013 cm is prevented from falling under gravity by the vertical field between two horizontal plates charged to a difference of potential of 8340 volts. The distance between the plates is 1.6 cm, and the density of oil is 920 kg m^{-3} . Calculate the magnitude of the charge on the drop ($g = 9.81 \text{ m s}^{-2}$). (O. & C.)

9. Two plane parallel conducting plates 1.50 cm apart are held horizontal, one above the other, in air. The upper plate is maintained at a positive potential of 1500 volts while the lower plate is earthed. Calculate the number of electrons which must be attached to a small oil drop of mass $4.90 \times 10^{-12} \text{ g}$, if it remains stationary in the air between the plates. (Assume that the density of air is negligible in comparison with that of oil.)

If the potential of the upper plate is suddenly changed to -1500 volts what is the initial acceleration of the charged drop? Indicate, giving reasons, how the acceleration will change.

10. Show how (i) the surface density, (ii) the intensity of electric field, (iii) the potential, varies over the surface of an elongated conductor charged with electricity. Describe experiments you would perform to support your answer in cases (i) and (iii).

Describe and explain the action of points on a charged conductor; and give two practical applications of the effect. (L.)

11. Describe carefully Faraday's ice-pail experiments and discuss the deductions to be drawn from them. How would you investigate experimentally the charge distribution over the surface of a conductor? (C.)

12. What is an *electric field*? With reference to such a field define *electric potential*.

Two plane parallel conducting plates are held horizontal, one above the other, in a vacuum. Electrons having a speed of $6.0 \times 10^8 \text{ cm s}^{-1}$ and moving normally

to the plates enter the region between them through a hole in the lower plate which is earthed. What potential must be applied to the other plate so that the electrons just fail to reach it? What is the subsequent motion of these electrons? Assume that the electrons do not interact with one another.

(Ratio of charge to mass of electron is 1.8×10^{11} coulomb kg^{-1} .) (N.)

13. Define the *electric potential* V and the *electric field strength* E at a point in an electrostatic field. How are they related? Write down an expression for the electric field strength at a point close to a charged conducting surface, in terms of the surface density of charge.

Corona discharge into the air from a charged conductor takes place when the potential gradient at its surface exceeds 3×10^6 volt metre $^{-1}$; a potential gradient of this magnitude also breaks down the insulation afforded by a solid dielectric. Calculate the greatest charge that can be placed on a conducting sphere of radius 20 cm supported in the atmosphere on a long insulating pillar; also calculate the corresponding potential of the sphere. Discuss whether this potential could be achieved if the pillar of insulating dielectric was only 50 cm long. (Take ϵ_0 to be 8.85×10^{-12} farad metre $^{-1}$.) (O.)

14. (i) A needle is mounted vertically, point upwards, on the plate (cap) of a gold leaf electroscope, the blunt end being in metallic contact with the plate. When a negatively charged body is brought close to the needle point, without touching it, and is then withdrawn the gold leaf is left with a permanent deflection. What is the sign of the charge causing this deflection, and how was this charge produced?

(ii) A gold leaf electroscope is so constructed that for a few degrees deflection the gold leaf touches the case and is thereby earthed. Describe and explain the behaviour of the leaf when :

- (a) a positively charged, insulated body is brought towards the plate (cap) of the electroscope until the leaf touches the case;
- (b) the positively charged body is then moved slowly closer to the plate;
- (c) the positively charged body is fixed close to the plate and the air between them is feebly ionized. (O. & C.)

chapter thirty-one

Capacitors

A capacitor (or 'condenser'), is a device for storing electricity. The earliest capacitor was invented—almost accidentally—by van Muschenbroek of Leyden, in about 1746, and became known as a Leyden jar. A present-day form of it is shown in Fig. 31.1 (i), J is a glass jar, FF are tin-foil coatings over the lower parts of its walls, and T is a knob connected to the inner coating. Modern forms of capacitor are shown at (ii) and (iv) in the figure. Essentially, all capacitors consist of two metal plates separated by an insulator. The insulator is called the dielectric; in some capacitors it is oil or air. Fig. 31.1 (iii) shows the conventional symbol for a capacitor.

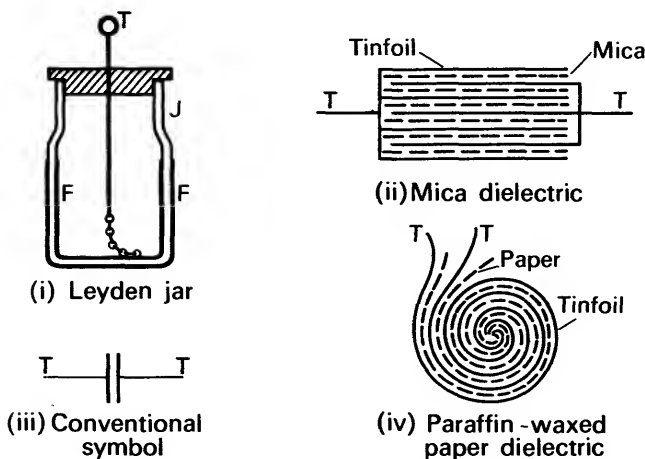


FIG. 31.1. Types of capacitor.

Charging a Capacitor

To study the action of a capacitor we need a paper one of about 4 microfarad capacitance (see later), a couple of high-tension batteries D and a high impedance voltmeter V such as a valve voltmeter reading to about 300 volts. We also need a two-way key (K in Fig. 31.2) and a poor conductor (R). The latter is a short stick of powdered and compressed carbon; it is called a radio resistor, and should have a resistance of about 5 megohms (p. 790). We connect the batteries in series, and measure their total voltage, V_0 , with the voltmeter (Fig. 31.2 (i)). We then connect up all the apparatus as shown in Fig. 31.2 (ii). If we close

the key at A, the capacitor is connected via the resistor to the battery, and the potential difference across the capacitor, V , which is measured

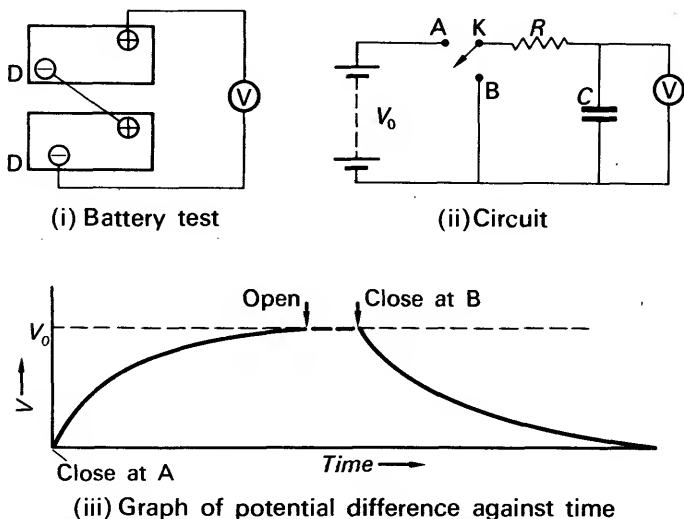


FIG. 31.2. Charging and discharging a capacitor through a resistor.

by the voltmeter, begins to rise (Fig. 31.2(iii)). The potential difference becomes steady when it is equal to the battery voltage V_0 . If we now open the key, the voltmeter reading stays unchanged (unless the capacitor is leaky). The capacitor is said to be charged, to the battery voltage; its condition does not depend at all on the resistor, whose only purpose was to slow down the charging process, so that we could follow it on the voltmeter.

Discharging a Capacitor

We can show that the charged capacitor is storing electricity by discharging it: if we put a piece of wire across its terminals, a fat spark passes just as the wire makes contact, and the voltmeter reading falls to zero.

If we now recharge the capacitor and then close the key at B, in Fig 31.2(ii), we allow the capacitor to discharge through the resistor R . The potential difference across it now falls to zero as slowly as it rose during charging.

Charging and Discharging Processes

When we connect a capacitor to a battery, electrons flow from the negative terminal of the battery on to the plate A of the capacitor

connected to it (Fig. 31.3); and, at the same rate, electrons flow from the other plate B of the capacitor towards the positive terminal of the battery. Positive and negative charges thus appear on the plates, and oppose the flow of electrons which causes them. As the charges accumulate, the potential difference between the plates increases, and the charging current falls to zero when the potential difference becomes equal to the battery voltage V_0 .

When the battery is disconnected and the plates are joined together by a wire, electrons flow back from plate A to plate B until the positive charge on B is completely neutralized. A current thus flows for a time in the wire, and at the end of the time the charges on the plates become zero.

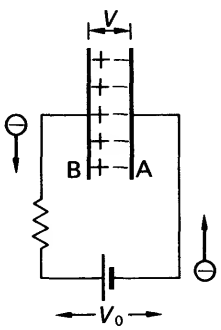


FIG. 31.3. A capacitor charging. (Resistance is shown because some is always present, even if only that of the connecting wires.)

Capacitors in A.C. Circuits

Capacitors are widely used in alternating current and radio circuits, because they can transmit alternating currents. To see how they do so, let us consider the circuit of Fig. 31.4, in which the capacitor may be connected across either of the batteries X, Y. When the key is closed at A, current flows from the battery X, and charges the plate D of the capacitor positively. If the key is now closed at B instead,

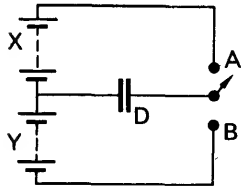


FIG. 31.4. Reversals of voltage applied to capacitor.

current flows from the battery Y; the plate D loses its positive charge and becomes negatively charged. Thus if the key is rocked rapidly between A and B, current surges backwards and forwards along the wires connected to the capacitor. An alternating voltage, as we shall see in Chapter 39, is one which reverses many times a second; when such a voltage is applied to a capacitor, therefore, an alternating current flows in the connecting wires.

Capacitance Definition, and Units

Experiments with a ballistic galvanometer, which measures quantity of electricity, show that, when a capacitor is charged to a potential difference V , the charges stored on its plates, $\pm Q$, are proportional to V . The ratio of the charge on either plate to the potential difference between the plates is called the *capacitance*, C , of the capacitor:

$$C = \frac{Q}{V}. \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Thus $Q = CV, \quad . \quad . \quad . \quad . \quad . \quad (2)$

and $V = \frac{Q}{C}. \quad . \quad . \quad . \quad . \quad . \quad (3)$

When Q is in coulombs (C) and V is in volts (V), then capacitance C is in farads (F). One farad (1F) is the capacitance of an extremely large capacitor. In practical circuits, such as in radio receivers, the capacitance of capacitors used are therefore expressed in *microfarads* (μF). One microfarad is one millionth part of a farad, that is, $1\mu\text{F} = 10^{-6}\text{F}$. It is also quite usual to express small capacitors, such as those used on radiograms for altering tone, in *picofarads* (pF). A picofarad is one millionth part of a microfarad, that is, $1\text{pF} = 10^{-6}\mu\text{F} = 10^{-12}\text{F}$.

Comparison of Capacitances

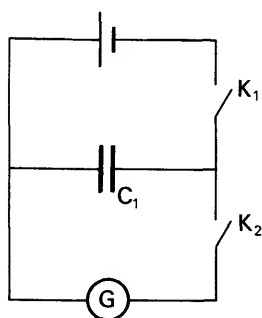


FIG. 31.5. Comparison of capacitances.

Large capacitances, of the order of microfarads, can be compared with the aid of a ballistic galvanometer. In this instrument, as explained on p. 920, the first 'throw' or deflection is proportional to the quantity of electricity discharged through it.

The circuit required is shown in Fig. 31.5. The capacitor of capacitance C_1 is charged by a battery of e.m.f. V , and then discharged through the ballistic galvanometer G . The corresponding first deflection θ_1 is observed. The capacitor is now replaced by another of capacitance C_2 , charged again by the battery, and the new deflection θ_2 is observed when the capacitor is discharged.

Now

$$\begin{aligned}\frac{Q_1}{Q_2} &= \frac{\theta_1}{\theta_2} \\ \therefore \frac{C_1 V}{C_2 V} &= \frac{\theta_1}{\theta_2} \\ \therefore \frac{C_1}{C_2} &= \frac{\theta_1}{\theta_2} \quad \quad \quad (4)\end{aligned}$$

If C_2 is a standard capacitor, whose value is known, then the capacitance of C_1 can be found.

Factors determining Capacitance. Variable Capacitor

We shall now find out by experiment what factors influence capacitance. To interpret our observations we shall require the formula for potential difference:

$$V = \frac{Q}{C}.$$

This shows that, when a capacitor is given a fixed charge, the potential difference between its plates is inversely proportional to its capacitance.

Distance between plates. In Fig. 31.6 (i), A and B are two metal plates, B being earthed, while A is insulated and connected to an electroscope.

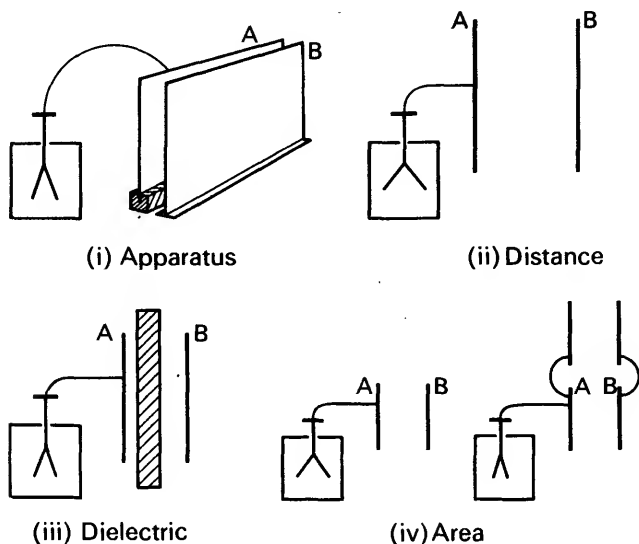


FIG. 31.6. Factors determining capacitance.

We set the plates close together, but not touching, and charge A from an electrophorus. The leaves of the electroscope diverge by an amount which measures the potential difference between the plates. If we move the plates further apart the leaves diverge further, showing that the potential difference has increased (Fig. 31.6(ii)). Since we have done nothing to increase the charge on the plates, the increase in potential difference must be due to a decrease in capacitance (see above equation). The capacitance of a capacitor therefore decreases when the separation of its plates is increased; we shall see in the next chapter that the capacitance is inversely proportional to the separation.

Dielectric. Let us now put a dielectric—a sheet of glass or ebonite—between the plates. The leaves diverge less, showing that the potential difference has decreased (Fig. 31.6(iii)). The capacitance has therefore increased, and the increase is due to the dielectric. By using several sheets of it we can show that the effect of the dielectric increases with its thickness. In practical capacitors the dielectric completely fills the space between the plates.

Area of plates. To see how the capacitance depends on the area of the plates, we set them at a known distance apart. We then take another pair of plates, at the same distance apart, and connect them to the first pair by wires held on insulating handles (Fig. 31.6(iv)). The leaves diverge less, showing that the capacitance has increased; it is, in fact, directly proportional to the area of the plates.

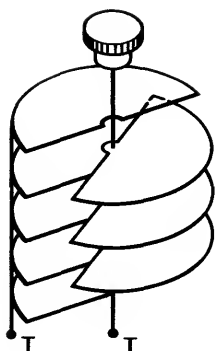


FIG. 31.7. Variable air capacitor.

A capacitor in which the effective area of the plates can be adjusted is called a *variable capacitor*. In the type shown in Fig. 31.7, the plates are semicircular and one set can be swung into or out of the other. The capacitance is proportional to the area of overlap of the plates. The plates are made of brass or aluminium, and the dielectric may be air, oil, or mica.

We shall see shortly that a capacitor with parallel plates, having a vacuum (or air, if we assume the permittivity of air is the same as a vacuum) between them, has a capacitance given by

$$C = \frac{\epsilon_0 A}{d}$$

where C = capacitance in farads (F), A = area of overlap of plates in m^2 , d = distance between plates in m and $\epsilon_0 = 8.854 \times 10^{-12}$ farad metre $^{-1}$.

If a material of permittivity ϵ completely fills the space between the plates, then the capacitance becomes:

$$C = \frac{\epsilon A}{d}$$

Capacitance Values. Isolated Sphere

Suppose a sphere of radius r metre situated in air is given a charge of Q coulombs. We assume, as on p. 747, that the charge on a sphere gives rise to potentials *on and outside* the sphere as if all the charge were concentrated at the *centre*. From p. 757, the surface of the sphere has a potential given by:

$$V = \frac{Q}{4\pi\epsilon_0 r}$$

$$\therefore \frac{Q}{V} = 4\pi\epsilon_0 r$$

$$\therefore \text{Capacitance, } C = 4\pi\epsilon_0 r \quad (1)$$

The other 'plate' of the capacitor is the earth.

Suppose $r = 10 \text{ cm} = 0.1 \text{ m}$. Then,

$$\begin{aligned} C &= 4\pi\epsilon_0 r = 4\pi \times 8.85 \times 10^{-12} \times 0.1 \text{ F} \\ &= 11 \times 10^{-12} \text{ F (approx.)} = 11 \text{ pF.} \end{aligned}$$

Concentric Spheres

Faraday used two concentric spheres to investigate the dielectric constant of liquids. Suppose a , b are the respective radii of the inner

and outer spheres (Fig. 31.8). Let $+Q$ be the charge given to the inner sphere and let the outer sphere be earthed, with air between them.

The induced charge on the outer sphere is $-Q$ (see p. 740). The potential V_a of the inner sphere = potential due to $+Q$ and potential due to $-Q = +\frac{Q}{4\pi\epsilon_0 a} - \frac{Q}{4\pi\epsilon_0 b}$, since the potential due to the charge $-Q$ is $-Q/4\pi\epsilon_0 b$ everywhere inside the larger sphere (see p. 757).

But $V_b = 0$, as the outer sphere is earthed.

$$\therefore \text{potential difference, } V = V_a - V_b = \frac{1}{4\pi\epsilon_0} \left(\frac{Q}{a} - \frac{Q}{b} \right)$$

$$\therefore V = \frac{Q}{4\pi\epsilon_0} \left(\frac{b-a}{ab} \right)$$

$$\therefore \frac{Q}{V} = \frac{4\pi\epsilon_0 ab}{b-a},$$

$$C = \frac{4\pi\epsilon_0 ab}{b-a} \quad (1)$$

As an example, suppose $b = 10$ cm and $a = 9$ cm

$$\begin{aligned} \therefore C &= \frac{4\pi\epsilon_0 ab}{b-a} \\ &= \frac{4\pi \times 8.85 \times 10^{-12} \times 0.1 \times 0.09}{(0.1 - 0.09)} \\ &= 100 \text{ pF (approx).} \end{aligned}$$

Note that the inclusion of a nearby second plate to the capacitor increases the capacitance. For an *isolated* sphere of radius 10 cm, the capacitance was 11 pF (p. 770).

Parallel Plate Capacitor

Suppose two parallel plates of a capacitor each have a charge numerically equal to Q , Fig. 31.9. The surface density σ is then Q/A

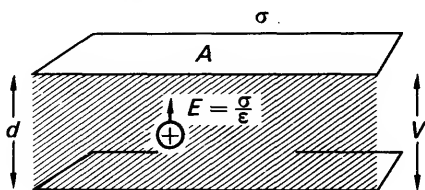


FIG. 31.9. Parallel-plate capacitor.

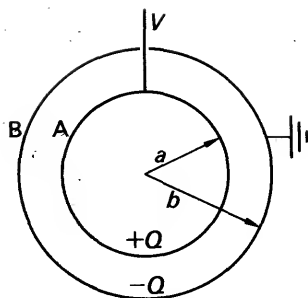


FIG. 31.8. Concentric spherical capacitor.

where A is the area of either plate, and the intensity between the plates, E , is given, from p. 749, by

$$E = \frac{\sigma}{\epsilon} = \frac{Q}{\epsilon A}$$

Now the work done in taking a unit charge from one plate to the other = force \times distance = $E \times d$ where d is the distance between the plates. But the work done per unit charge = V , the p.d. between the plates.

$$\therefore V = \frac{\sigma}{\epsilon} d = \frac{Q}{\epsilon A} d$$

$$\therefore \frac{Q}{V} = \frac{\epsilon A}{d}$$

$$\therefore C = \frac{\epsilon A}{d} \quad (1)$$

It should be noted that this formula for C is approximate, as the field becomes non-uniform at the edges. See Fig. 30.30, p. 755.

Dielectric Constant (Relative Permittivity) and Strength

To study the effect of the dielectric in a capacitor, the capacitance of a given capacitor must be measured: first without a dielectric, and then with one. We shall see later how this can be done (p. 776).

The ratio of the capacitance with and without the dielectric between the plates is called the *dielectric constant* (or *relative permittivity*) of the material used. The expression 'without a dielectric' strictly means 'with the plates in a vacuum'; but the effect of air on the capacitance of a capacitor is so small that for most purposes it may be neglected. The dielectric constant of a substance is denoted by the letter ϵ : thus

$$\epsilon = \frac{\text{capacitance of given capacitor, with space between plates filled with dielectric}}{\text{capacitance of same capacitor with plates in vacuo}}$$

If we take the case of a parallel plate capacitor as an example, then

$$\epsilon_r = \frac{\epsilon A/d}{\epsilon_0 A/d} = \frac{\epsilon}{\epsilon_0}$$

Thus the dielectric constant is the ratio of the permittivity of the substance to that of free space. It is for this reason that dielectric constant is also known as 'relative permittivity'. Note that the dielectric constant is a pure number and has no dimensions, unlike ϵ or ϵ_0 .

The following table gives the value of dielectric constant, and also of *dielectric strength*, for various substances. The strength of a dielectric is the potential gradient at which its insulation breaks down, and a spark passes through it. A solid dielectric is ruined by such a breakdown, but a liquid or gaseous one heals up as soon as the applied potential difference is reduced.

Water is not suitable for a dielectric in practice, because it is a good

insulator only when it is very pure, and to remove all matter dissolved in it is almost impossible.

PROPERTIES OF DIELECTRICS

Substance	Dielectric constant	Dielectric strength, kilovolts per mm
Glass	5-10	30-150
Mica	6	80-200
Ebonite	2-8	30-110
Ice*	94	—
Paraffin wax	2	15-50
Paraffined paper	2	40-60
Paraffin oil	4-7	—
Methyl alcohol*	32	—
Water*	81	—
Air	1-0005	—
Sulphur dioxide*	1-01	—

* Polar molecules (see p. 774).

Action of Dielectric

The explanation of dielectric action which we shall now give is similar in principle to Faraday's, but expressed in modern terms—there was no knowledge of electrons in his day.

We regard a molecule as a collection of atomic nuclei, positively charged, and surrounded by a cloud of negative electrons. When a dielectric is in a charged capacitor, its molecules are in an electric field; the nuclei are urged in the direction of the field, and the electrons in the opposite direction (Fig. 31.10 (i)). Thus each molecule is distorted, or polarized: one end has an excess of positive charge, the other an excess of negative. At the surfaces of the dielectric, therefore, charges appear, as shown in Fig. 31.10 (ii). These charges are of opposite sign to the charges on the plates, and so reduce the potential difference between the plates.

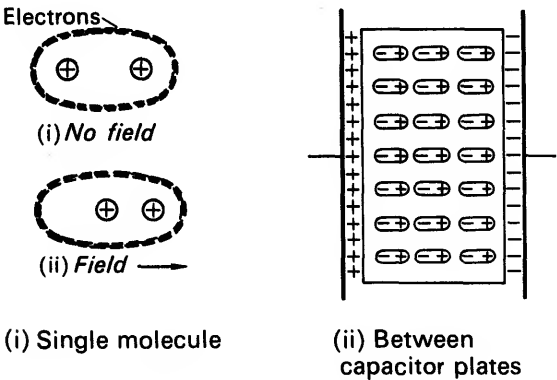


FIG. 31.10. Polarization of a dielectric.

If the capacitor is connected to a battery, then its potential difference is constant; but the surface charges on the dielectric still increase its capacitance. They do so because they offset the mutual repulsions of the charges on the plates, and so enable greater charges to accumulate there before the potential difference rises to the battery voltage.

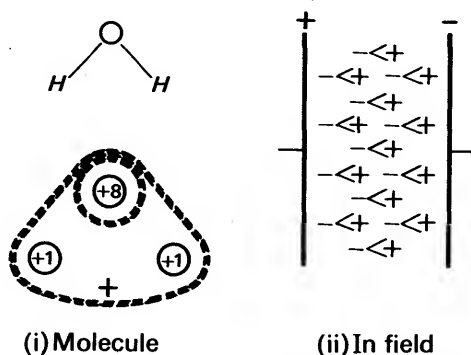


FIG. 31.11. Water as a dielectric.

Some molecules, we believe, are permanently polarized: they are called *polar molecules*. For example, the water molecule consists of an oxygen atom, O , with two hydrogen atoms H , making roughly a right-angled structure (Fig. 31.11 (i)). Oxygen has a nuclear charge of $+8e$, where e is the electronic charge, and has eight electrons. Hydrogen has a nuclear charge of $+e$, and one electron. In the water molecule, the two electrons from the hydrogen atom move in paths which surround the oxygen nucleus. Thus they are partly added to the oxygen atom, and partly withdrawn from the hydrogen atoms. On the average, therefore, the apex of the triangle is negatively charged, and its base is positively charged. In an electric field, water molecules tend to orient themselves as shown in Fig. 31.11 (ii). The effect of this, in a capacitor, is to increase the capacitance in the way already described. The increase is, in fact, much greater than that obtained with a dielectric which is polarized merely by the action of the field.

ϵ_0 and its measurement

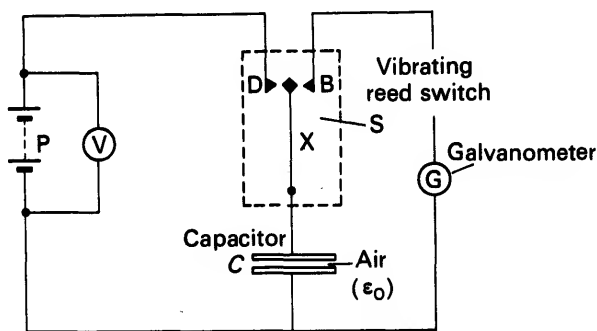
We can now see how the units of ϵ_0 may be stated in a more convenient manner and how its magnitude may be measured.

Units. From $C = \frac{\epsilon_0 A}{d}$, we have $\epsilon_0 = \frac{Cd}{A}$

Thus the unit of $\epsilon_0 = \frac{\text{farad} \times \text{metre}}{\text{metre}^2}$

= farad metre⁻¹ (see also p. 743).

Measurement. In order to find the magnitude of ϵ_0 , the circuit in Fig. 31.12 is used.

FIG. 31.12. Measurement of ϵ_0 .

C is a parallel plate capacitor, which may be made of sheets of glass or perspex coated with aluminium foil or aquadag. The two conducting surfaces are placed facing inwards, so that only air is present between these plates. The area A of the plates in metre^2 , and the separation d in metres, are measured. P is a high tension supply capable of delivering about 200 V, and G is a calibrated sensitive galvanometer, such as a 'Scalamp' type. S is a *vibrating switch* unit, energized by a low a.c. voltage from the mains. When operating, the vibrating bar X touches and then B , and the motion is repeated at the mains frequency, fifty times a second. When the switch is in contact with D , the capacitor is charged from the supply P to a potential difference of V volt, measured on the voltmeter. When the contact moves over to B , the capacitor discharges through the galvanometer. The galvanometer thus receives fifty pulses of charge per second. This gives an average steady reading on the galvanometer, corresponding to a mean current I .

Now from previous, $C = \epsilon_0 A/d$. Thus on charging, the charge stored, Q , is given by

$$Q = CV = \frac{\epsilon_0 VA}{d}.$$

The capacitor is discharged fifty times per second. Since the current is the charge flowing per second,

$$\therefore I = \frac{\epsilon_0 VA \cdot 50}{d} \text{ ampere}$$

$$\therefore \epsilon_0 = \frac{Id}{50VA} \text{ farad metre}^{-1}$$

The following results were obtained in one experiment:

$$A = 0.0317 \text{ m}^2, d = 1.0 \text{ cm} = 0.010 \text{ m}, V = 150 \text{ V}, I = 0.21 \times 10^{-6} \text{ A}$$

$$\begin{aligned} \therefore \epsilon_0 &= \frac{Id}{50VA} \\ &= \frac{0.21 \times 10^{-6} \times 0.01}{50 \times 150 \times 0.0317} \\ &= 8.8 \times 10^{-12} \text{ farad metre}^{-1} \end{aligned}$$

As very small currents are concerned, care must be taken to make

the apparatus of high quality insulating material, otherwise leakage currents will lead to serious error.

This method can also be used to find the permittivity of various materials. Thus if the experiment is repeated with a material of permittivity ϵ completely filling the space between the plates, then $\epsilon = I'd/50 \text{ VA}$, where I' is the new current.

If only the relative permittivity or dielectric constant, ϵ_r , is required, there is no need to know the p.d. supplied or the dimensions of the capacitor. In this case,

$$\epsilon_r = \frac{\epsilon}{\epsilon_0} = \frac{I'd/50 \text{ VA}}{Id/50 \text{ VA}} = \frac{I'}{I}$$

Thus ϵ_r is the ratio of the respective currents in G with and without the dielectric between the plates.

Measurement of Capacitance

If a standard capacitor is available, an unknown capacitance can be measured as described on p. 768.

If no standard capacitor is available, the method of the last section can be employed. The unknown capacitor replaces C in the circuit of Fig. 31.12. The current I in the galvanometer, and the p.d. supplied, V , are then measured. Now the charge $Q = CV$. This is discharged fifty times per second. Since the current is the charge flowing per second,

$$\therefore I = 50 CV$$

$$\therefore C = \frac{I}{50 V} \text{ farad.}$$

Arrangements of Capacitors

In radio circuits, capacitors often appear in arrangements whose resultant capacitances must be known. To derive expressions for these, we need the equation defining capacitance in its three possible forms:

$$C = \frac{Q}{V}, \quad V = \frac{Q}{C}, \quad Q = CV.$$

In Parallel. Fig. 31.13 shows three capacitors, having all their left-hand plates connected together, and all their right-hand plates likewise.

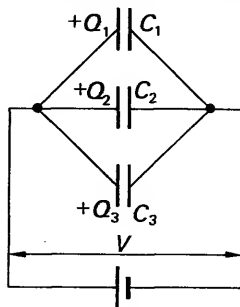


FIG. 31.13. Capacitors in parallel.

They are said to be connected in parallel. If a cell is now connected across them, they all have the same potential difference V . (For, if they had not, current would flow from one to another until they had.) The charges on the individual capacitors are respectively

$$\begin{aligned} Q_1 &= C_1 V \\ Q_2 &= C_2 V \\ Q_3 &= C_3 V \end{aligned} \quad (1)$$

The total charge on the system of capacitors is

$$Q = Q_1 + Q_2 + Q_3 = (C_1 + C_2 + C_3)V.$$

And the system is therefore equivalent to a single capacitor, of capacitance

$$C = \frac{Q}{V} = C_1 + C_2 + C_3.$$

Thus when capacitors are connected in parallel, their resultant capacitance is the sum of their individual capacitances. It is greater than the greatest individual one.

In Series. Fig. 31.14 shows three capacitors having the right-hand plate of one connected to the left-hand of the next, and so on—connected in series. When a cell is connected across the ends of the system, a charge Q is transferred from the plate H to the plate A, a charge $-Q$ being left on H. This charge induces a charge $+Q$ on plate G; similarly, charges

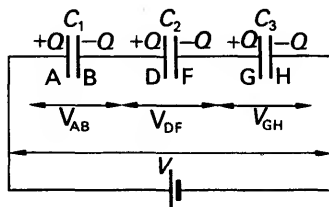


FIG. 31.14. Capacitors in series.

appear on all the other capacitor plates, as shown in the figure. (The induced and inducing charges are equal because the capacitor plates are very large and very close together, in effect, either may be said to enclose the other.) The potential differences across the individual capacitors are, therefore, given by

$$\left. \begin{aligned} V_{AB} &= \frac{Q}{C_1} \\ V_{DF} &= \frac{Q}{C_2} \\ V_{GH} &= \frac{Q}{C_3} \end{aligned} \right\} \quad (2)$$

The sum of these is equal to the applied potential difference V because the work done in taking a unit charge from H to A is the sum of the work done in taking it from H to G, from F to D, and from B to A. Therefore

$$V = V_{AB} + V_{DF} + V_{GH}$$

$$= Q \left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} \right). \quad (3)$$

The resultant capacitance of the system is the ratio of the charge stored to the applied potential difference, V . The charge stored is equal to Q , because, if the battery is removed, and the plates HA joined by a wire, a charge Q will pass through that wire, and the whole system will be discharged. The resultant capacitance is therefore given by

$$C = \frac{Q}{V}, \quad \text{or} \quad \frac{1}{C} = \frac{V}{Q},$$

whence, by equation (3),

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}. \quad (4)$$

Thus, to find the resultant capacitance of capacitors in series, we must add their reciprocals of the individual capacitances. The resultant is less than the smallest individual.

Comparison of Series and Parallel Arrangements. Let us compare Figs. 31.13 and 31.14. In Fig. 31.14, where the capacitors are in series, all the capacitors carry the same charge, which is equal to the charge carried by the system as a whole, Q . The potential difference applied to the system, however, is divided amongst the capacitors, in inverse proportion to their capacitances (equations (2)). In Fig. 31.13, where the capacitors are in parallel, they all have the same potential difference; but the charge stored is divided amongst them, in direct proportion to the capacitances (equations (1)).

EXAMPLE

Find the charges on the capacitors in Fig. 31.15, and the potential differences across them.

Capacitance between A and B,

$$C' = C_2 + C_3 = 3 \mu\text{F}.$$

Overall capacitance A to D,

$$C = \frac{C_1 C'}{C_1 + C'} = \frac{2 \times 3}{2 + 3} = 1.2 \mu\text{F}.$$

Charge stored in this

$$= Q_1 = Q_2 + Q_3 = CV = 1.2 \times 10^{-6} \times 120$$

$$= 144 \times 10^{-6} \text{ coulomb},$$

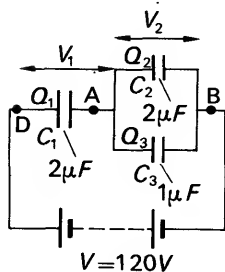


FIG. 31.15. Example.

$$\begin{aligned}\therefore V_1 &= \frac{Q_1}{C_1} = \frac{144 \times 10^{-6}}{2 \times 10^{-6}} = 72 \text{ volt,} \\ V_2 &= V - V_1 = 120 - 72 = 48 \text{ volt,} \\ Q_2 &= C_2 V_2 = 2 \times 10^{-6} \times 48 = 96 \times 10^{-6} \text{ coulomb,} \\ Q_3 &= C_3 V_2 = 10^{-6} \times 48 = 48 \times 10^{-6} \text{ coulomb.}\end{aligned}$$

Energy of a Charged Capacitor

A charged capacitor is a store of electrical energy, as we may see from the vigorous spark it can give on discharge. To find the energy stored, let us suppose that the capacitor, of capacitance C , is already charged to a potential difference V . And let us suppose that we wish to increase the charge on its plates from Q to $Q + \delta Q$, where δQ is very small. Then we must transfer a charge δQ from the negative plate to the positive. In doing so, we shall increase the potential difference by the amount

$$\delta V = \frac{\delta Q}{C}.$$

But if δQ is very small compared with Q , δV will be very small compared with V , and the potential difference will be almost constant at the value V . Then the work done in displacing the charge δQ will be

$$\delta W = V \delta Q.$$

from the definition of potential difference. But

$$V = \frac{Q}{C},$$

and therefore

$$\delta W = \frac{Q}{C} \delta Q.$$

If we now suppose that the capacitor is at first completely discharged, and the charged until the final charge on the plates has some definite value Q_1 , then the work done in charging it is

$$\int_{Q=0}^{Q=Q_1} dW = \int_0^{Q_1} \frac{Q dQ}{C} = \left[\frac{1}{2} \frac{Q^2}{C} \right]_0^{Q_1} = \frac{1}{2} \frac{Q_1^2}{C}.$$

In general, therefore, the energy stored by a capacitor of capacitance C , carrying a charge Q , at a potential difference V , is

$$\begin{aligned}W &= \frac{1}{2} (Q^2 / C) \\ &= \frac{1}{2} QV \\ &= \frac{1}{2} CV^2\end{aligned} \quad (1)$$

If C is measured in farad, Q in coulomb and V in volt, then the expressions derived in (1) will give the energy W in joules.

When the spheres are connected by a fine wire, the potentials become equalized.

Suppose this is V . Then, since $Q = CV$ and the total charge is constant,

original total charge = total final charge.

$$\therefore 4\pi\epsilon_0(5 \times 10^{-2} \times 600 + 10 \times 10^{-2} \times 300) = 4\pi\epsilon_0(5 \times 10^{-2} V + 10 \times 10^{-2} V).$$

Solving,

$$V = 400 \text{ volt}$$

$$\begin{aligned} \therefore \text{total final energy} &= \frac{1}{2}C_1V^2 + \frac{1}{2}C_2V^2 = \frac{1}{2}(C_1 + C_2)V^2 \\ &= \frac{1}{2} \times \frac{1}{9 \times 10^9} (5 \times 10^{-2} + 10 \times 10^{-2}) \times 400^2 \\ &= 13\frac{1}{3} \times 10^{-7} \text{ J.} \end{aligned} \quad \text{(ii)}$$

From (i) and (ii), $13\frac{1}{3} \times 10^{-7} \text{ J}$ of electrical energy have been converted into heat. This occurred because a current flows in the connecting wire when the two spheres are connected.

3. Define *electrical capacitance*. Describe experiments to demonstrate the factors which determine its value for a parallel plate capacitance.

The plates of a parallel plate air capacitor consisting of two circular plates, each of 10 cm radius, placed 2 mm apart, are connected to the terminals of an electrostatic voltmeter. The system is charged to give a reading of 100 on the voltmeter scale. The space between the plates is then filled with oil of dielectric constant 4.7 and the voltmeter reading falls to 25. Calculate the capacitance of the voltmeter. You may assume that the voltage recorded by the voltmeter is proportional to the scale reading (N).

Suppose V is the initial p.d. across the air capacitor and voltmeter, and let C_1 be the voltmeter capacitance.

$$\text{Then} \quad \text{total charge} = CV + C_1V = (C + C_1)V. \quad \text{(i)}$$

When the plates are filled with oil the capacitance increases to $4.7C$, and the p.d. falls to V_1 . But the total charge remains constant.

$$\therefore 4.7CV_1 + C_1V_1 = (C + C_1)V, \quad \text{from (i).}$$

$$\therefore (4.7C + C_1)V_1 = (C + C_1)V,$$

$$\therefore \frac{4.7C + C_1}{C + C_1} = \frac{V}{V_1} = \frac{100}{25} = 4,$$

$$\therefore 0.7C = 3C_1,$$

$$\therefore C_1 = \frac{0.7C}{3} = \frac{7}{30}C.$$

Now $C = \epsilon_0 A/d$, where A is in metre^2 and d is in metres.

$$\therefore C = \frac{8.85 \times 10^{-12} \times \pi \times (10 \times 10^{-2})^2}{2 \times 10^{-3}} \text{ F}$$

$$= 1.4 \times 10^{-10} \text{ F (approx.).}$$

$$\therefore C_1 = \frac{7}{30} \times 1.4 \times 10^{-10} \text{ F} = 3.3 \times 10^{-11} \text{ F.}$$

EXERCISES 31

1. Derive an expression for the capacitance of a parallel plate capacitor. Describe an experiment to demonstrate the variation of capacitance with *one* of the physical quantities involved.

A $2.5\ \mu\text{F}$ capacitor is charged to a potential difference of 100 volts and is disconnected from the supply. Its terminals are then connected to those of an uncharged $10\ \mu\text{F}$ capacitor. Find (a) the resulting potential difference across the two capacitors and (b) the total energy stored in them. Compare the result in (b) with the energy originally stored in the $2.5\ \mu\text{F}$ capacitor and comment on the difference. (L.)

2. Explain the meaning of the term *capacitance* as used in electrostatics.

A potential difference of 90 V is applied across uncharged capacitors of $2\ \mu\text{F}$, $3\ \mu\text{F}$ and $1.5\ \mu\text{F}$ connected in series. Across which capacitor is the potential difference least? Explain this and find the numerical value of this potential difference. (N.)

3. Define *capacitance* of a conductor. State the factors on which its value depends and describe simple experiments to justify your answers.

The circular plates, *A* and *B*, of a parallel plate air capacitor have each an effective diameter of 10.0 cm and are 2.0 mm apart. The plates, *C* and *D*, of a similar capacitor have each an effective diameter of 12.0 cm and are 3.0 mm apart. *A* is earthed, *B* and *C* are connected together and *D* is connected to the positive pole of a 120-volt battery whose negative pole is earthed. Calculate (a) the combined capacitance of the arrangement, (b) the energy stored in it, (c) the energy stored in each capacitor. (L.)

4. Explain what is meant by *dielectric constant* (*relative permittivity*). State two physical properties desirable in a material to be used as the dielectric in a capacitor.

A sheet of paper 4.0 cm wide and 1.5×10^{-3} cm thick between metal foil of the same width is used to make a $2.0\ \mu\text{F}$ capacitor. If the dielectric constant (relative permittivity) of the paper is 2.5, what length of paper is required? ($\epsilon_0 = 8.85 \times 10^{-12}$ farad m^{-1} .) (N.)

5. Define *capacitance*. Obtain from first principles a formula for the capacitance of a parallel-plate capacitor.

Outline, without full experimental details, a method of determining the relative permittivity (dielectric constant) of a dielectric material.

A sensitive moving-coil meter with a suspended system that can swing freely is calibrated both as a ballistic galvanometer and as a galvanometer to record steady currents. Its sensitivity is 8×10^{-8} coulombs per division ballistically, and 2.5×10^{-7} amperes per division for steady currents. When a capacitor is charged from a 24 volt battery and discharged through the meter, the ballistic throw is 30 divisions. When a 2 volt battery is used, and the capacitor is charged and discharged through the meter many times per second by means of a vibrating-reed switch, the steady deflection is 80 divisions. Find (a) the value of the capacitance, and (b) the number of times the capacitor is discharged per second. (O.)

6. Explain what is meant by the capacitance of a capacitor and define a unit in which it is measured.

Explain the principle of the guard ring capacitor. How could it be used to pass a charge of known value through a galvanometer?

A voltmeter of high internal resistance is connected across the terminals of a $10\ \mu\text{F}$ capacitor which is initially charged to 410 V. When the charging source is removed the voltmeter reading falls and reaches 390 V after 10.0 seconds.

Calculate the internal resistance of the voltmeter. (You may use an approximate method of calculation, but must explain why it is approximate.) (O. & C.)

7. Derive an expression for the energy stored in a capacitor C when there is a potential difference V between the plates. If C is in microfarads and V is in volts, express the result in joules.

Show that when a battery is used to charge a capacitor through a resistor, the heat dissipated in the circuit is equal to the energy stored in the capacitor.

Describe the structure of a 1 microfarad capacitor and describe an experiment to compare the capacitances of two capacitors of this type. (N.)

8. Define *electrostatic potential* and *capacitance*. Assuming that the electric intensity between the plates of a parallel-plate air capacitor is $1/\epsilon_0$ times the charge density on one of the plates, obtain an expression for its capacitance. If you were provided with a capacitor of unknown capacitance and a standard capacitor of comparable capacitance, how would you find the value of the unknown capacitance? A parallel-plate air capacitor is charged to a potential difference of 300 volts and is then connected in parallel with another capacitor of equal dimensions with ebonite as dielectric. The potential of the combination is found to be 75 volts. Calculate the dielectric constant of the ebonite. (N.)

9. Deduce expressions for the combined capacitance of two capacitors (a) connected in series, (b) connected in parallel. Describe how *one* of these expressions may be verified by experiment.

A fixed capacitor of capacitance 10^{-4} microfarad is connected in series with a variable capacitor the capacitance of which may be varied from zero to 10^{-4} microfarad in steps of 10^{-6} microfarad. These two in series are connected in parallel with a third capacitor of capacitance 5×10^{-4} microfarad. Calculate (i) the maximum capacitance of the whole combination, (ii) the smallest change in this capacitance which can be produced by the arrangement. (N.)

10. Explain how the strength of the electric field at any point is related to the electric potential at and near the point.

A parallel plate capacitor consists of two large plates 2 cm apart, the dielectric on one side of the middle plane between the plates consisting of air, and on the other side of an insulating material of dielectric constant 4.2. Calculate the strength of the electric field in the half occupied by air, if the difference of potential between the plates is 500 volts. (L.)

11. Define *capacitance* (or capacity), *relative permittivity* (or dielectric constant), microfarad (μF).

Derive an approximate expression for the capacitance of an air capacitor consisting of two parallel opposite circular plates of radius r cm at a distance t cm apart. Explain on what the degree of approximation depends.

Two capacitors of capacitances respectively $2 \mu\text{F}$ and $3 \mu\text{F}$ are joined in series between points A and B. What capacitance must be placed in parallel with the $2 \mu\text{F}$ capacitor in order to increase the capacitance from A to B by $0.8 \mu\text{F}$? (L.)

12. State the law of force between electric charges. Write down expressions for (a) the electric field strength, and (b) the electric potential, at a point in air situated at a distance r from a point charge Q . State the units in which each of the quantities is measured.

Describe the Faraday ice-pail experiment. Illustrate the successive steps in the experiment by a series of simple diagrams, and state the conclusions that can be drawn from the experiment.

The sensitivity of a ballistic galvanometer is found by charging a 0.001 microfarad capacitor from a 12-volt battery, and discharging it through the meter,

which gives a corrected swing of 16.8 divisions. A large hollow metal can is then connected to one terminal of the galvanometer, while the other is earthed; an insulated conducting sphere, of 5 cm radius, is charged to an unknown potential V , thrust right inside the can, and allowed to touch its inner surface; as a result, the galvanometer gives a corrected swing of 21.0 divisions. Find, in coulombs, the charge on the sphere, and also the value in V in volts. (O.)

13. Define potential, capacitance.

Obtain from first principles a formula for the capacitance of a parallel-plate capacitor.

The plates of such a capacitor are each 0.4 m square, and separated by 10^{-3} m, the space between being filled with a medium of relative permittivity 5. A vibrating contact, with frequency 50 seconds⁻¹, repeatedly connects the capacitor across a 120-volt battery and then discharges it through a galvanometer whose resistance is of the order of 50 ohm. Calculate the current recorded, and explain why this is independent of the actual value of the galvanometer resistance. (Take the permittivity of space to be 8.85×10^{-12} F m⁻¹.) (O.)

14. Define 'the capacitance of a capacitor' and show how it leads to a practical unit of capacitance. Why is it necessary to specify 'capacitor' in this definition?

State the physical factors which affect the magnitude of a capacitance, and describe an experiment to demonstrate the effect of one of them.

A charged 20 microfarad capacitor A is connected to an electrostatic voltmeter of infinite resistance and negligible capacitance which reads 500 volts. A 0.25 microfarad capacitor B is now connected in parallel with A. The capacitor B is then disconnected from A, discharged, and the process of connection and discharge repeated ten times in all. Calculate (a) what fraction of the charge on A remains after the first connexion, (b) what fraction remains eventually, and (c) the final reading of the voltmeter (C.)

chapter thirty-two

Current Electricity

OHM'S AND JOULE'S LAWS: RESISTANCE AND POWER

Discovery and Electric Current

By the middle of the eighteenth century, electrostatics was a well-established branch of physics. Machines had been invented which could produce by friction great amounts of charge, giving sparks and electric shocks. The momentary current (as we would now call it) carried by the spark or the body was called a 'discharge'.

In 1786 Galvani, while dissecting a frog, noticed that its leg-muscle twitched when one of his assistants produced an electric spark in another part of the room. He also found that, when a frog's leg-muscle was hung by a copper hook from an iron stand, the muscle twitched whenever it swung so as to touch the stand. Galvani supposed that the electricity which caused the twitching was generated within the muscle, but his fellow-Italian Volta believed that it arose from the contact of the two different metals. Volta turned out to be right, and

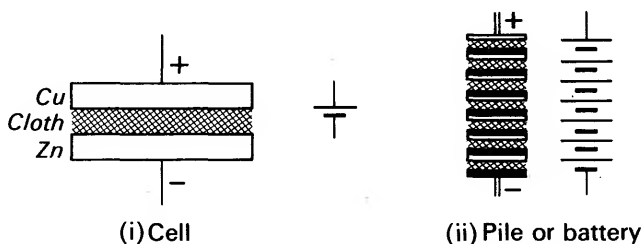


FIG. 32.1. Voltaic cell and pile, with conventional symbols.

in 1799 he discovered how to obtain from two metals a continuous supply of electricity: he placed a piece of cloth soaked in brine between copper and zinc plates (Fig. 32.1 (i)). The arrangement is called a *voltaic cell*, and the metal plates its 'poles'; the copper is known as the positive pole, the zinc as the negative cell. Volta increased the power by building a pile of cells, with the zinc of one cell resting on the copper of the other (Fig. 32.1 (ii)). From this pile he obtained sparks and shocks similar to those given by electrostatic machines.

Shortly after, it was found that water was decomposed into hydrogen and oxygen when connected to a voltaic pile. This was the earliest discovery of the chemical effect of an electric current. The heating

effect was also soon found, but the magnetic effect, the most important effect, was discovered some twenty years later.

Ohm's Law

The properties of an electric circuit, as distinct from the effects of a current, were first studied by Ohm in 1826. He set out to find how the length of wire in a circuit affected the current through it—in modern language, he investigated electrical resistance. In his first experiments he used voltaic piles as sources of current, but he found that the current which they gave fluctuated considerably, and he later replaced them by thermocouples (p. 803). He passed the currents through various lengths of brass wire, 0.037 cm in diameter, and observed them on a torsion balance galvanometer; this is represented by G in Fig. 32.2 (i). No unit of current had been defined at the time of these experiments, but physicists had agreed to take the strength of a current as proportional to its magnetic field. Ohm found that the

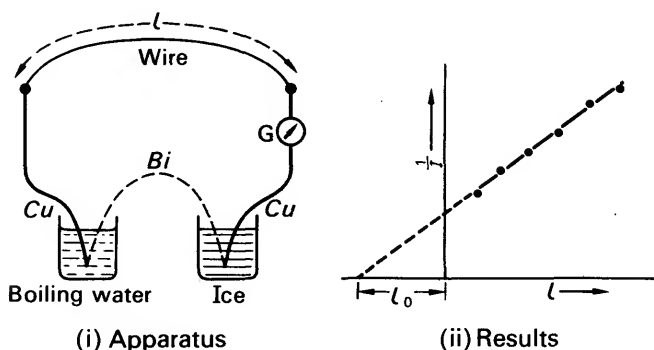


FIG. 32.2. Ohm's experiment.

current I in his experiments was almost inversely proportional to the length of wire, l , in the circuit. He plotted the reciprocal of the current (in arbitrary units) against the length l , and got a straight line, as shown in Fig. 32.2 (ii). Thus

$$I \propto \frac{1}{l_0 + l},$$

where l_0 is the intercept of the line on the axis of length. Ohm explained this result by supposing, naturally, that the thermocouples and galvanometer, as well as the wire, offered resistance to the current. He interpreted the constant l_0 as the length of wire equal in resistance to the galvanometer and thermocouples. The *ohm*, symbol Ω , is the unit of electrical resistance (see p. 790).

Mechanism of Metallic Conduction

The conduction of electricity in metals is due to free electrons. Free electrons have thermal energy, and wander randomly through the

metal from atom to atom. When a battery is connected across the ends of the metal, an electric field is set up. The electrons are now accelerated by the field, so they gain velocity and energy. When they 'collide' with an atom vibrating about its fixed mean position (called a 'lattice site'), they give up some of their energy to it. The amplitude of the vibrations is then increased and the temperature of the metal rises. The electrons are then again accelerated by the field and again give up some energy. Although their movement is erratic, on the average the electrons drift in the direction of the field with a mean speed we calculate shortly. This drift constitutes an 'electric current'. It will be noted that heat is generated by the collision of electrons whichever way they flow. Thus the heating effect of a current—called *Joule heating* (p. 790)—is irreversible, that is, it still occurs when the current in a wire is reversed.

A simple calculation enables the average drift speed to be estimated. Fig. 32.3 shows a portion of a copper wire of cross-sectional area A

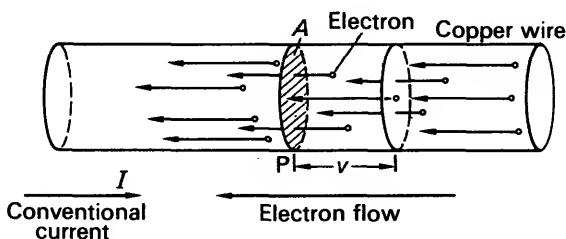


FIG. 32.3. Theory of metallic conduction.

through which a current I is flowing. We suppose that there are n electrons per unit volume, and that each electron carries a charge e . Now in one second all those electrons within a distance v to the right of the plane at P, that is, in a volume Av , will flow through this plane, as shown. This volume contains nAv electrons and hence a charge $nAve$. Thus a charge of $nAve$ per second passes P, and so the current I is given by

$$I = nAve \quad (1)$$

To find the order of magnitude of v , suppose $I = 10\text{A}$, $A = 1\text{ mm}^2 = 10^{-6}\text{m}^2$, $e = 1.6 \times 10^{-19}\text{C}$, and $n = 10^{28}$ electrons m^{-3} . Then, from (1),

$$\begin{aligned} v &= \frac{I}{nAe} = \frac{10}{10^{28} \times 10^{-6} \times 1.6 \times 10^{-19}} \\ &= \frac{1}{160} \text{ m s}^{-1} \text{ (approx.)} \end{aligned}$$

This is a surprisingly slow drift compared with the average thermal speeds, which are of the order of several hundred metres per second (p. 234).

Resistivity

Ohm showed, by using wires of different length and diameter, that the resistance of a wire, R , is proportional to its length, l , and inversely proportional to its cross-sectional area A . The truth of this can easily be demonstrated today by experiments with a Wheatstone bridge (see p. 829) and suitable lengths of wire. We have, then, for a given wire,

$$R \propto \frac{l}{A};$$

we may therefore write

$$R = \rho \frac{l}{A}, \quad (1)$$

where ρ is a constant for the material of the wire. It is called the *resistivity* of that material.

To define it in words, we imagine a rectangular prism of the material, of unit length and unit cross-section. Then $l = 1$, $A = 1$, and $R = \rho$. Thus the resistivity of a substance is the resistance between the faces of a rectangular prism of the substance, which is 1 cm long and whose cross-sectional area is 1 cm². One unit of resistivity is 1 ohm cm, because, from equation (1),

$$\rho = R \frac{A}{l}, \quad (2)$$

which has the units

$$\text{ohms} \times \frac{\text{cm}^2}{\text{cm}} = \text{ohms} \times \text{cm}.$$

Resistivities are thus often expressed in microhm centimetres; 1 microhm = 10^{-6} ohm. The SI unit is the *ohm metre* (Ω m). Using it, R is in ohms when l is in metres and A is in metre² in equation (2).

RESISTIVITIES

Substance	Resistivity ρ , ohm m (at 20°C)	Temperature coefficient α , K ⁻¹
Aluminium	2.82×10^{-8}	0.0039
Brass	c. 8×10^{-8}	c. 0.0015
Constantan ¹	c. 49×10^{-8}	0.00001
Copper	1.72×10^{-8}	0.0043
Iron	c. 9.8×10^{-8}	0.0056
Manganin ²	c. 44×10^{-8}	c. 0.00001
Mercury	95.77×10^{-8}	0.00091
Nichrome ³	c. 100×10^{-8}	0.0004
Silver	1.62×10^{-8}	0.0039
Tungsten ⁴	5.5×10^{-8}	0.0058
Carbon (graphite) . . .	33 to 185×10^{-8}	-0.0006 to -0.0012

¹ Also called Eureka; 60 per cent Cu, 40 per cent Ni.

² 84 per cent Cu, 12 per cent Mn, 4 per cent Ni; used for resistance boxes and shunts.

³ Ni-Cu-Cr; used for electric fires—does not oxidize at 1000°C.

⁴ Used for lamp filaments—melts at 3380°C.

The resistivity of a metal is increased by even small amounts of impurity; and alloys, such as Constantan, may have resistivities far greater than any of their constituents.

Ohm's Theory of the Circuit

We have seen that Ohm abandoned the use of voltaic piles in his experiments, because the currents which they gave were not steady. He attributed this to fluctuations in their 'exciting force'—electromotive force, as we now call it. Similarly, when he used thermocouples, he found that the current through a given circuit increased when the difference in temperature between the couples was increased. He was thus led to propose a 'mathematical law of the galvanic circuit':

$$I = \frac{E}{R} \quad (3)$$

Here I stands for the strength of the current, E for the exciting force (e.m.f.), and R for the total resistance of the circuit.

Demonstration of Ohm's Law

Ohm showed that his law applied not only to a complete circuit but to any part of it. To understand this, let us consider an experiment which can easily be done with modern apparatus. As shown in Fig. 32.4 (a), we connect in series the following apparatus:

- (i) one or more accumulators, S ;
- (ii) a milliammeter reading to 15 milliamperes;
- (iii) a wire-wound resistor R of the order of 50 ohms;
- (iv) a suitable variable resistance or *rheostat* P of the same order of resistance.

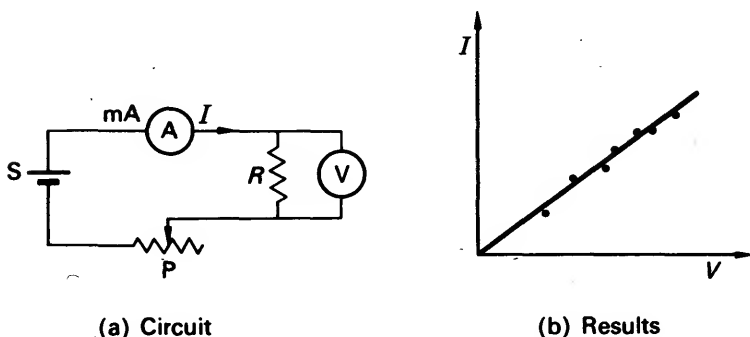


FIG. 32.4. Demonstration of Ohm's law.

Across the resistor R we connect a voltmeter to measure the potential difference V across R . This must be a voltmeter such as a potentiometer (p. 817) whose calibration does not depend on Ohm's law, otherwise the experiment would not be valid. The milliammeter calibration likewise must not depend on Ohm's law. By adjusting the resistor P we vary the current I through the circuit, and at each value of I we

measure V . On plotting V against I we get a straight line through the origin, as in Fig. 32.4 (b); this shows that the potential difference across the resistor R is proportional to the current through it:

$$V \propto I. \quad (4)$$

Thus, taking into account that the resistance of a conductor depends on its temperature and on other physical conditions such as mechanical strain, Ohm's law can be stated as follows:

Under constant physical conditions, the potential difference across a conductor is proportional to the current through it.

The law is obeyed by the most important class of conductors—metals—and by some others, such as carbon. It is not obeyed by some crystals, such as silicon carbide, nor by some conducting solutions, nor by diode valves, nor—as in a neon lamp—by gases.

Resistance

From Ohm's law, it follows that

$$\frac{V}{I} = R, \text{ a constant.} \quad (5)$$

R is *defined* as the 'resistance' of the conductor.

The unit of potential difference, V , is the *volt*, symbol V ; the unit of current, I , is the *ampere*, symbol A ; the unit of resistance, R , is the *ohm*, symbol Ω . The ohm is thus the resistance of a conductor through which a current of one ampere flows when a potential difference (p.d.) of one volt is across it.

From the above equation, it also follows that

$$V = IR, \text{ and } I = \frac{V}{R}. \quad (6)$$

Smaller units of current are the milliampere (one-thousandth of an ampere), symbol mA and the micro-ampere (one-millionth of an ampere), symbol μA . Smaller units of p.d. are the millivolt ($1/1000$ V) and the microvolt ($1/10^6$ V). A small unit of resistance is the microhm ($1/10^6$ ohm); larger units are the kilohm ($1,000$ ohms) and the megohm (10^6 ohms).

HEAT AND POWER

Electrical Heating. Joule's Laws

In 1841 Joule studied the heating effect of an electric current by passing it through a coil of wire in a jar of water (Fig. 32.5). He used various currents, measured by an early form of galvanometer G , and various lengths of wire, but always the same mass of water. The rise in temperature of the water, in a given time, was then proportional to the heat developed by the current in that time. Joule found that the heat produced in a given time, with a given wire, was proportional to

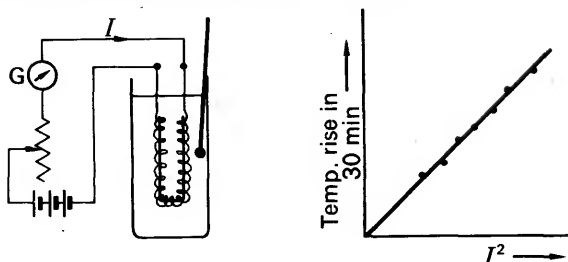


FIG. 32.5. Joule's experiment on heating effect of current.

I^2 , where I is the current flowing. If H is the heat produced per second, then

$$H \propto I^2. \quad (7)$$

Joule also made experiments on the heat produced by a given current in different wires. He used wires of different lengths, but of the same diameter, and of the same material; he found that the rate at which heat was produced, by a given current, was proportional to the length of the wire. That is to say, he found that the rate of heat production was proportional to what Ohm had already called the resistance of the wire:

$$H \propto R. \quad (8)$$

Relationships (7) and (8) together give

$$H \propto I^2 R. \quad (9)$$

Mechanism of the Heating Effect

Heat is a form of energy. The heat produced per second by a current in a wire is therefore a measure of the energy which it liberates in one second, as it flows through the wire. The heat is produced, we suppose, by the free electrons as they move through the metal. On their way they collide frequently with atoms; at each collision they lose some of their kinetic energy, and give it to the atoms which they strike. Thus, as the current flows through the wire, it increases the kinetic energy of vibration of the metal atoms: it generates heat in the wire. The electrical resistance of the metal is due, we say, to its atoms obstructing the drift of the electrons past them: it is analogous to mechanical friction. As the current flows through the wire, the energy lost per second by the electrons is the electrical power supplied by the battery which maintains the current. That power comes, as we shall see later, from the chemical energy liberated by these actions within the battery.

Potential Difference and Energy

On p. 750 we defined the potential difference V_{AB} between two points, A and B, as the work done by an external agent in taking a unit

positive charge from B to A (Fig. 32.6 (i)). This definition applies equally well to points in an electrostatic field and to points on a conductor carrying a current.

In Fig. 32.6 (ii), D represents any electrical device or circuit element: a lamp, motor, or battery on charge, for example. A current of I amperes flows through it from the terminal A to the terminal B; if it flows for t seconds, the charge Q which it carries from A to B is, since a current is the quantity of electricity per second flowing,

$$Q = It \text{ coulombs.} \quad (10)$$

Let us suppose that the device D liberates a total amount of energy W joules in the time t ; this total may be made up of heat, light, sound mechanical work, chemical transformation, and any other forms of energy. Then W is the amount of electrical energy given up by the charge Q in passing through the device D from A to B.

$$\therefore W = QV_{AB} \quad (11)$$

where V_{AB} is the potential difference between A and B in volts.

The work, in all its forms, which the current I does in t seconds as it flows through the device, is therefore

$$W = IV_{AB}t, \quad (12)$$

by equations (10) and (11).

Electrical Power

The energy liberated per second in the device is defined as its electrical power. The electrical power, P , supplied is given, from above, by

$$P = \frac{W}{t} = \frac{IV_{AB}t}{t}$$

or

$$P = IV_{AB} \quad (13)$$

When an electric current flows through a wire or 'passive' resistor, all the power which it conveys to the wire appears as heat. If I is the current, R is the resistance, then $V_{AB} = IR$, Fig. 32.7.

$$\therefore P = I^2R. \quad (14)$$

Also,

$$P = \frac{V_{AB}^2}{R}. \quad (15)$$

The power, P , is in watts (W) when I is in amp, R is in ohms, and V_{AB} is in volts. 1 kilowatt (kW) = 1000 watts.

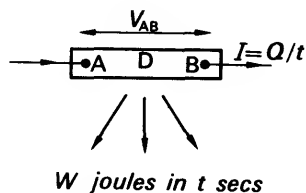
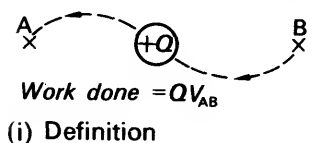


FIG. 32.6. Potential difference and energy.

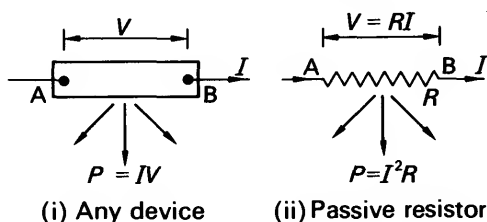


FIG. 32.7. Power equations.

The formulae for power, $P = I^2 R$ or V^2/R , is true only when all the electrical power supplied is dissipated as heat. As we shall see, the formulae do not hold when part of the electrical energy supplied is converted into mechanical work, as in a motor, or into chemical energy, as in an accumulator being charged. A device which converts all the electrical energy supplied to it into heat is called a 'passive' resistor; it may be a wire, or a strip of carbon, or a liquid which conducts electricity but is not decomposed by it. Since the joule (J) is the unit of heat, it follows that, for a resistor, the heat H in it in joules is given by

$$H = IVt$$

$$\text{or by} \quad H = I^2 R t \quad . \quad . \quad . \quad (16)$$

$$\text{or by} \quad H = \frac{V^2 t}{R}$$

The units of I , V , R are amperes (A), volts (V), ohms (Ω) respectively.

High-tension Transmission

When electricity has to be transmitted from a source, such as a power station, to a distant load, such as a factory, the two must be connected by cables. These cables have resistance, which is in effect added to the internal resistance of the generator; power is wasted in them as heat. If r is the total resistance of the cables, and I the supply current, the power wasted is $I^2 r$. The power delivered to the factory is IV , where V is the potential difference at the factory. Economy requires the waste power, $I^2 r$, to be small; but it also requires the cables to be thin, and therefore cheap to buy and erect. The thinner the cables, however, the higher their resistance r . Thus the most economical way to transmit the power is to make the current, I , as small as possible; this means making the potential difference V as high as possible. When large amounts of power are to be transmitted, therefore, very high voltages are used: 132000 volts on the main lines of the British grid, 6000 volts on subsidiary lines. These voltages are much too high to be brought into a house, or even a factory. They are stepped down by transformers, in a way which we shall describe later; stepping-down in that way is possible only with alternating current, which is one of the main reasons why alternating current is so widely used.

EXAMPLE

An electric heating element to dissipate 480 watts on 240 V mains is to be made from Nichrome ribbon 1 mm wide and thickness 0.05 mm. Calculate the length of ribbon required if the resistivity of Nichrome is 1.1×10^{-6} ohm metre.

Power, $P = \frac{V^2}{R}$

$$\therefore R = \frac{V^2}{P} = \frac{240^2}{480} = 120\Omega$$

The area A of cross-section of the ribbon = $1 \times 0.05 \text{ mm}^2 = 0.05 \times 10^{-6} \text{ m}^2$.

From $R = \frac{\rho l}{A}$

$$\therefore l = \frac{R \cdot A}{\rho} = \frac{120 \times 0.05 \times 10^{-6}}{1.1 \times 10^{-6}} = 5.45 \text{ metre}$$

Summary of Formulae Related to Power and Ohm's Law

In any device whatever (Fig. 32.7 (i)):

Electrical power consumed = power developed in other forms,

$$P = IV,$$

$$\text{watts} = \text{amperes} \times \text{volts.}$$

In a passive resistor (Fig. 32.7 (ii)):

$$(i) \quad V = IR; \quad I = \frac{V}{R}; \quad R = \frac{V}{I},$$

$$\text{volts} = \text{ohms} \times \text{amperes.}$$

(ii) Power consumed = heat developed per second, in watts.

$$P = I^2 R = IV = \frac{V^2}{R}.$$

(iii) Heat developed in time t :

Electrical energy consumed = heat developed in joules

$$I^2 R t = IVt = \frac{V^2}{R} t.$$

Board of Trade (commercial) unit = kilowatt hour (kWh) = kilowatt \times hour

$$= 3.6 \times 10^6 \text{ joule.}$$

ELECTROMOTIVE FORCE**E.M.F. Internal Resistance**

If we take a high-tension battery, and connect a high resistance volt-meter across it, the meter reads about 120 volts. Across two batteries

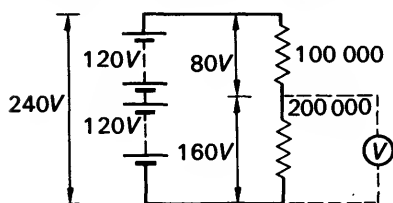


FIG. 32.8. P.D. with moving-coil voltmeter.

resistor, we should find that the potential difference across all three added up to the same value. It appears that the batteries always maintain a total potential difference of 240 volts across any circuit to which they are joined. This constant potential difference represents what Ohm called the 'exciting force' of the batteries. Since it is the property which enables the batteries to maintain a flow of electricity in a circuit, we may call it their electromotive force.

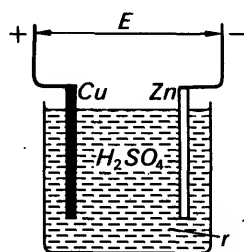
Now let us connect a lower resistance across the batteries of Fig. 32.8—say 10000 ohms. We find the potential difference across their terminals falls slightly. And if we use a still lower resistance—say 1000 ohms—then the potential difference falls greatly, perhaps to about $\frac{2}{3}$ of its open-circuit value.

We suppose, therefore, that the batteries have some internal resistance—the resistance r of the chemical solutions between the plates (Fig. 32.9 (i)). This is analogous to the resistance of the wires of Ohm's thermocouples (p. 780). When an appreciable current I flows from the battery, it sets up a potential difference rI across the internal resistance, and by that amount makes the external potential difference less than the electromotive force, E (Fig. 32.9 (ii)). In a rough-and-ready way we may represent the battery as a source of constant potential difference E , in series with a passive resistor r (Fig. 32.9 (iii)).

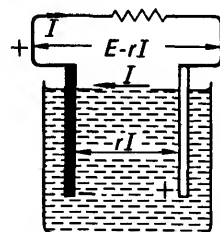
Electromotive Force and Energy

To get a rigorous definition of electromotive force, let us first imagine that we pass a current through the device of Fig. 32.10 in opposition to its e.m.f. We can do this by connecting its terminals AC, via a resistance

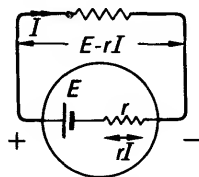
in series it reads 240 volts. Let us now connect the two batteries in series with two resistors, of resistance 200000 and 100000 ohms, as in Fig. 32.8. By using the voltmeter, we find that the potential difference across the 200000-ohm resistor is 160 volts, and across the 100000-ohm resistor 80 volts. These add up to 240 volts; if we inserted a third



(i) A voltaic cell on open circuit



(ii) On load



(iii) Representation

FIG. 32.9. E.m.f. and internal resistance.

R , to a battery D which has a greater e.m.f. (Fig. 32.10 (i)). If a charge Q passes round the circuit in a given time, then the work done in carrying it from A to B , against the potential difference E , is QE joules. This work appears as chemical changes in the source of E . Now suppose that we remove the battery D , and connect a resistor across the terminals AC (Fig. 32.10 (ii)). The potential difference will now send a current round the circuit in the opposite direction to the previous current. And when a

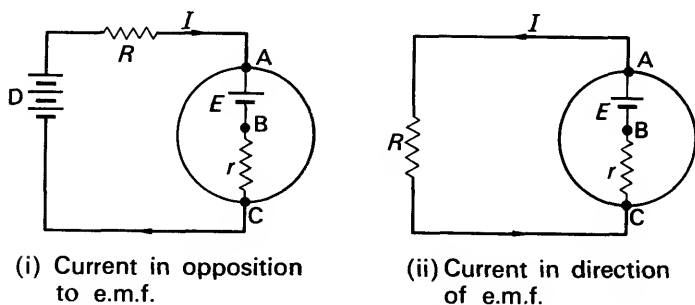


FIG. 32.10. Illustrating definition of e.m.f.

charge Q has passed, the energy delivered by the source of e.m.f. E will be QE joules. The chemical changes in the source will have been reversed, and will have given up this amount of energy, as electrical energy. The current, in passing round the circuit, will have converted this energy into heat. Some of the heat will have been dissipated in the external resistance R , some in the internal resistance r .

Our picture of a source of current as a constant potential difference in series with a resistance is over-simplified, but it has brought us to the point where we can make a definition of e.m.f. which is both rigorous and intelligible. We shall make it first in terms of charge, later in terms of current.

In terms of charge: if a device has an electromotive force E , then, in passing a charge Q round a circuit joined to it, it liberates an amount of electrical energy equal to QE . If a charge Q is passed through the source against its e.m.f., then the work done against the e.m.f. is QE . The above definition of e.m.f. does not depend on any assumptions about the nature of its source.

If a device of e.m.f. E passes a steady current I for a time t , then the charge that it circulates is

$$Q = It.$$

Thus:

$$\text{electrical energy liberated, } W, = QE = IEt, \quad (17)$$

and

$$\text{electrical power generated, } P, = \frac{W}{t} = EI. \quad (18)$$

We can now define e.m.f. in terms of power and current, and therefore

in a way suitable for dealing with circuit problems. From equation (18)

$$P = EI,$$

or

$$E = \frac{P}{I}.$$

Thus the *e.m.f. of a device is the ratio of the electrical power which it generates, to the current which it delivers*. If current is forced through a device in opposition to its *e.m.f.*, then equation (18) gives the power consumed in overcoming the *e.m.f.*

Electromotive force resembles potential difference in that both can be defined as the ratio of power to current. The unit of *e.m.f.* is therefore 1 watt per ampere, or 1 volt; and the *e.m.f. of a source, in volts, is numerically equal to the power which it generates when it delivers a current of 1 ampere*.

Representation of an E.M.F. and Internal Resistance

Sources of *e.m.f.* differ widely in their nature. In a thermocouple, the *e.m.f.* arises at the junction of the two metals—this point is sometimes called the seat of the *e.m.f.* In a voltaic cell, we believe, the *e.m.f.* arises at the interfaces of the plates and solutions—part at one interface, the rest at the other. In each of these sources we can distinguish between the seat, or seats, of the *e.m.f.*, and that of the internal resistance, which is in the bulk of the solution or the wires of the thermocouple. The *e.m.f.* of a dynamo, however, does not arise at a point: it acts along the wires of the armature coil as they move in the magnetic field (p. 915). Here we cannot distinguish between the seats of the *e.m.f.* and the internal resistance. In solving circuit problems, however, it is helpful to show the *e.m.f.* and internal resistance separately, although as a rule they are physically inextricable.

Ohm's Law for Complete Circuit

Fig. 32.11 shows a source of current connected to a passive resistor—called the load—of resistance R . To find the current I , we equate the power generated by the source to the heat developed per second in the resistances:

$$EI = I^2r + I^2R.$$

Thus

$$E = Ir + IR, \quad (19)$$

whence

$$I = \frac{E}{R + r}. \quad (20)$$

Equation (20) asserts that the current is equal to the *e.m.f.* of the source divided by the total resistance of the circuit; it is Ohm's original statement of his law (p. 786). Equation (19) asserts that the sum of the potential differences across the resistances is equal to the *e.m.f.* The potential difference Ir appears across the internal resistance, and is often called the voltage drop; because of it, the

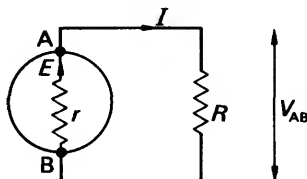


FIG. 32.11. A complete circuit.

potential difference between the terminals of the cell, V_{AB} , falls when the current taken, I , is increased:

$$V_{AB} = E - Ir. \quad (21)$$

Terminal Potential Difference

The quantity V_{AB} is often called the terminal potential difference of the cell; it is also the potential difference across the load. Equations (19) and (21) give

$$V_{AB} = IR, \quad (22)$$

which is Ohm's law for the load alone; we could have written it down directly. Equations (20) and (22) together give the terminal potential difference in terms of the e.m.f. and the resistances:

$$V_{AB} = IR = \frac{E}{R+r}R. \quad (23)$$

Output and Efficiency

The power delivered to the load in Fig. 32.10 is called the output power, P_{out} ; its value is

$$\begin{aligned} P_{out} &= IV_{AB} \\ &= I^2R. \end{aligned} \quad (24)$$

The power generated by the source of current is

$$P_{gen} = IE. \quad (25)$$

The difference between the power generated and the output is the power wasted as heat in the source: I^2r . The ratio of the power output to the power generated is the efficiency, η , of the circuit as a whole:

$$\eta = \frac{P_{out}}{P_{gen}}. \quad (26)$$

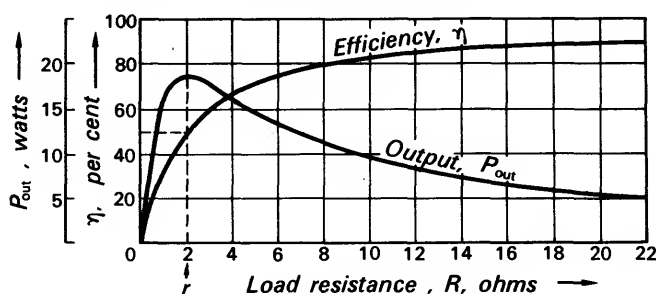
By equations (24) and (25), therefore,

$$\begin{aligned} \eta &= \frac{P_{out}}{P_{gen}} = \frac{IV_{AB}}{IE} \\ &= \frac{V_{AB}}{E}. \end{aligned}$$

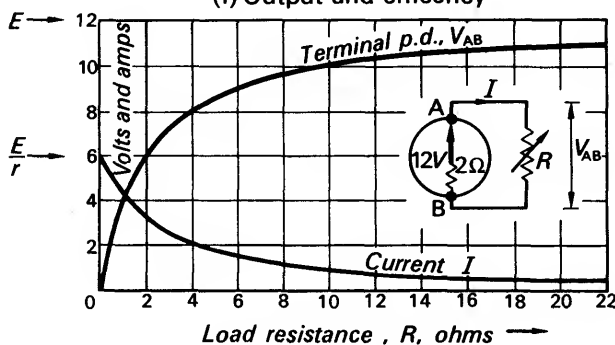
Equation (23) now gives

$$\eta = \frac{R}{R+r}. \quad (27)$$

This shows that the efficiency tends to unity (or 100 per cent) as the load resistance R tends to infinity. For high efficiency the load resistance must be several times the internal resistance of the source. When the load resistance is equal to the internal resistance, the efficiency is 50 per cent. (See Fig. 32.12 (i).)



(i) Output and efficiency



(ii) Current and potential difference

FIG. 32.12. Effect of varying load resistance.

Power Variation

Now let us consider how the power output varies with the load resistance. Equations (24) and (20) give

$$P_{out} = I^2 R,$$

and

$$I = \frac{E}{R + r},$$

whence

$$P_{out} = \frac{E^2 R}{(R + r)^2}.$$

If we take fixed values of E and r , and plot P_{out} as a function of R , we find that it passes through a maximum when $R = r$ (Fig. 32.12 (i)). We can get the same result in a more general way by differentiating P_{out} with respect to R , and equating the differential coefficient to zero. Physically, this result means that the power output is very small when R is either very large or very small, compared with r . When R is very large, the terminal potential difference, V_{AB} , approaches a constant value equal to the e.m.f. E (Fig. 32.12 (ii)); as R is increased the current I falls, and the power IV_{AB} falls with it. When R is very small, the current approaches the constant value E/r , but the potential difference (which is equal to IR) falls steadily with R ; the power output therefore

falls likewise. Consequently the power output is greatest for a moderate value of R ; the mathematics show that this value is actually $R = r$.

To prove $R = r$, differentiate the expression for P_{out} given on p. 799, with respect to R . Then

$$E^2 \frac{(R+r)^2 - R \cdot 2(R+r)}{(R+r)^4} = 0, \text{ for a maximum.}$$

From the numerator, $r^2 - R^2 = 0$, or $R = r$.

Examples of Loads in Electrical Circuits

The loading on a dynamo or battery is generally adjusted for high efficiency, because that means greatest economy. Also, if a large dynamo were used with a load not much greater than its internal resistance, the current would be so large that the heat generated in the internal resistance would ruin the machine. With batteries and dynamos, therefore, the load resistance is made many times greater than the internal resistance.

Loading for greatest power output is common in communication engineering. For example, the last transistor in a receiver delivers electrical power to the loudspeaker, which the speaker converts into mechanical power as sound-waves (p. 596). Because it converts electrical energy into mechanical energy, and not heat, the loudspeaker is not a passive resistor, and the simple equations above do not apply to it. Nevertheless, circuit conditions can be specified which enable the transistor to deliver the greatest power to the speaker; these are similar to the condition of equal load and internal resistances, and are usually satisfied in practice.

Load not a Passive Resistor

As an example of a load which is not a passive resistor, we shall take an accumulator being charged. The charging is done by connecting the accumulator X in opposition to a source of greater e.m.f., Y in Fig. 32.13,

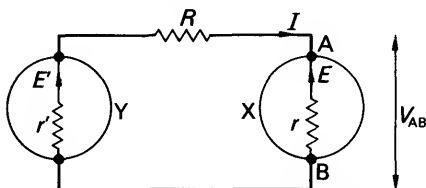


FIG. 32.13. Accumulator charging.

via a controlling resistor R . If E , E' and r , r' are the e.m.f. and internal resistances of X and Y respectively, then the current I is given by the equation:

$$\left. \begin{array}{l} \text{power generated} \\ \text{in } Y \end{array} \right\} = \left\{ \begin{array}{l} \text{power converted to} \\ \text{chemical energy} \\ \text{in } X \end{array} \right\} + \left\{ \begin{array}{l} \text{power dissipated} \\ \text{as heat in all} \\ \text{resistances} \end{array} \right.$$

$$E'I = EI + I^2R + I^2r' + I^2r. \quad (28)$$

Thus $(E' - E)I = I^2(R + r' + r),$

whence $I = \frac{E' - E}{R + r' + r}. \quad (29)$

The potential difference across the accumulator, V_{AB} , is given by

$$\left. \begin{array}{c} \text{power delivered} \\ \text{to X} \end{array} \right\} = \left\{ \begin{array}{c} \text{power converted to} \\ \text{chemical energy} \end{array} \right\} + \left\{ \begin{array}{c} \text{power dissipated} \\ \text{as heat} \end{array} \right\}$$

$$IV_{AB} = IE + I^2r.$$

Hence $V_{AB} = E + Ir. \quad (30)$

Equation (30) shows that, when current is driven through a generator in opposition to its e.m.f., then the potential difference across the generator is equal to the *sum* of its e.m.f. and the voltage drop across its internal resistance. This result follows at once from energy considerations, as we have just seen.

Cells in Series and Parallel

When cells or batteries are in series and assist each other, then the total e.m.f.

$$E = E_1 + E_2 + E_3 + \dots, \quad (31)$$

and the total internal resistance

$$r = r_1 + r_2 + r_3 + \dots, \quad (32)$$

where E_1, E_2 are the individual e.m.f.s and r_1, r_2 are the corresponding internal resistances. If one cell, e.m.f. E_2 say, is turned round 'in opposition' to the others, then $E = E_1 - E_2 + E_3 + \dots$; but the total internal resistance remains unaltered.

When *similar* cells are in parallel, the total e.m.f. = E , the e.m.f. of any one of them. The internal resistance r is here given by

$$\frac{1}{r} = \frac{1}{r_1} + \frac{1}{r_1} + \dots, \quad (33)$$

where r_1 is the internal resistance of each cell. If different cells are in parallel, there is no simple formula for the total e.m.f. and the total internal resistance, and any calculations involving circuits with such cells are dealt with by applying Kirchhoff's laws (see p. 827).

Summary of Formulae Involving E.M.F.

(i) *Any load*:

$$(a) \left. \begin{array}{c} \text{power generated} \end{array} \right\} = \left\{ \begin{array}{c} \text{power supplied} \\ \text{to load} \end{array} \right\} + \left\{ \begin{array}{c} \text{power dissipated} \\ \text{in internal} \\ \text{resistance} \end{array} \right\}$$

$$EI = IV_{AB} + I^2r,$$

(b) terminal p.d. = e.m.f. - voltage drop in internal resistance

$$V_{AB} = E - rI.$$

(ii) *Passive resistance load:*

power equation: $EI = I^2R + I^2r,$

current: $I = \frac{E}{R+r},$

terminal p.d.: $V_{AB} = E - rI = RI$
 $= E \frac{R}{R+r}.$

EXAMPLES

1. What is meant by the *electromotive force* of a cell?

A voltmeter is connected in parallel with a variable resistance, R , which is in series with an ammeter and a cell. For one value of R the meters read 0.3 amp and 0.9 volt. For another value of R the readings are 0.25 amp and 1.0 volt. Find the values of R , the e.m.f. of the cell, and the internal resistance of the cell. What assumptions are made about the resistance of the meters in the calculation?

If in this experiment the ammeter had a resistance of 10 ohms and the voltmeter a resistance of 100 ohms and R was 2 ohms, what would the meters read? (*L.*)

First part (see p. 796).

Second part. The voltmeter reads the p.d. across the cell if the resistances of the meters are neglected. Thus, with the usual notation,

$$E - Ir = 0.9, \text{ or } E - 0.3r = 0.9 \quad . \quad . \quad . \quad (i)$$

and $E - 0.25r = 1.0. \quad . \quad . \quad . \quad (ii)$

Subtracting (i) from (ii),

$$0.05r = 0.1, \text{ i.e. } r = 2 \text{ ohms.}$$

Also, from (i),

$$E = 0.3r + 0.9 = 0.6 + 0.9 = 1.5 \text{ volts.}$$

Further, $R_1 = \frac{V}{I} = \frac{0.9}{0.3} = 3 \text{ ohms.}$

and $R_2 = \frac{1.0}{0.25} = 4 \text{ ohms.}$

If the voltmeter has 100 ohms resistance and is in parallel with the 2 ohms resistance, the combined resistance R is given by

$$\frac{1}{R} = \frac{1}{2} + \frac{1}{100} = \frac{51}{100} \quad \text{or} \quad R = \frac{100}{51} \text{ ohms.}$$

$$\therefore \text{ current, } I = \frac{E}{\text{Total resistance}} \\ = \frac{1.5}{\frac{100}{51} + 10 + 2} = 0.11 \text{ A.}$$

$$\text{Also, voltmeter reading} = IR = 0.11 \times \frac{100}{51} = 0.21 \text{ volt.}$$

2. Define *internal resistance* of a voltaic cell. Describe one method of finding by experiment the internal resistance of a primary cell.

Two Daniell cells A and B are connected in series with a coil of resistance 9.8 ohms. A voltmeter of very high resistance connected to the terminals of A reads 0.96 volt and when connected to the terminals of B it reads 1.00 volt. Find the internal resistance of each cell. (Take the e.m.f. of a Daniell cell as 1.08 volts.) (L.)

First part. The 'internal resistance' is the resistance of the chemicals inside the cell between the poles, and is given by (drop in terminal p.d.)/current, when the cell is used. The potentiometer may be used to measure the internal resistance. See p. 820.

Second part. The p.d. across both cells = $0.96 + 1.00 = 1.96$ volts
= p.d. across 9.8 ohms.

$$\therefore \text{current flowing, } I, = \frac{V}{R} = \frac{1.96}{9.8} = 0.2 \text{ A.}$$

Now terminal p.d. across each cell = $E - Ir$.

$$\therefore \text{for cell A, } 0.96 = 1.08 - 0.2r, \quad \text{or} \quad r = 0.6 \text{ ohm.}$$

$$\text{for cell B, } 1.00 = 1.08 - 0.2r, \quad \text{or} \quad r = 0.4 \text{ ohm.}$$

THE THERMOELECTRIC EFFECT

Seebeck Effect

The heating effect of the current converts electrical energy into heat, but we have not so far described any mechanism which converts heat into electrical energy. This was discovered by Seebeck in 1822. In his experiments he connected a plate of bismuth between copper wires leading to a galvanometer, as shown in Fig. 32.14 (i). He found that if

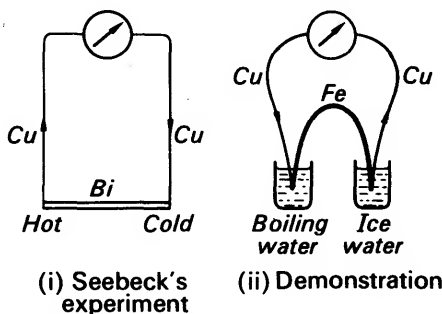


FIG. 32.14. The thermo-electric effect.

one of the bismuth-copper junctions was heated, while the other was kept cool, then a current flowed through the galvanometer. The direction of the current was from the copper to the bismuth at the cold junction. We can easily repeat Seebeck's experiment, using copper and iron wires and a galvanometer capable of indicating a few microamperes (p. 882) (Fig. 32.14 (ii)).

Thermocouples

Seebeck went on to show that a current flowed, without a battery, in any circuit containing two different metals, with their two junctions

at different temperatures. Currents obtained in this way are called thermo-electric currents, and a pair of metals, with their junctions at different temperatures, are said to form a thermocouple. The following is a list of metals, such that if any two of them form a thermocouple, then the current will flow from the higher to the lower in the list, across the cold junction :

Antimony, Iron, Zinc, Lead, Copper, Platinum, Bismuth.

Thermo-electric currents often appear when they are not wanted; they may arise from small differences in purity of two samples of the same metal, and from small differences of temperature—due, perhaps, to the warmth of the hand. They can cause a great deal of trouble in circuits used for precise measurements, or for detecting other small currents, not of thermal origin. As sources of electrical energy, thermo-electric currents are neither convenient nor economical, but they have been used—in gas-driven radio sets. Their only wide application is in the measurement of temperature, and of other quantities, such as radiant energy, which can be measured by a temperature rise.

Variation of Thermoelectric E.M.F. with Temperature

On p. 825 we shall see how thermo-electric e.m.f.s are measured. When the cold junction of a given thermocouple is kept constant at 0°C , and

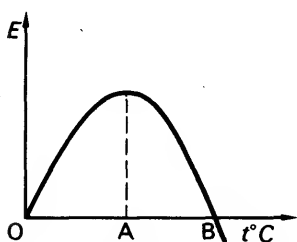


FIG. 32.15. Thermo-electric e.m.f. variation with temperature.

the hot junction temperature $t^{\circ}\text{C}$ is varied, the e.m.f. E is found to vary as $E = at + bt^2$, where a, b are constants. This is a parabola-shaped curve (Fig. 32.15). The temperature A corresponding to the maximum e.m.f. is known as the *neutral temperature*; it is about 250°C for a copper-iron thermocouple. Beyond the temperature B , known as the *inversion temperature*, the e.m.f. reverses. Thermo-electric thermo-

meters, which utilize thermocouples, are used only as far as the neutral temperature, as the same e.m.f. is obtained at two different temperatures, from Fig. 32.15.

Peltier and Thomson Effects

When a current flows along the junction A of two metals in series, heat is evolved or absorbed at A depending on the current direction. This is known as the *Peltier effect*. It has no connexion with the usual heating or Joule effect of a current, discussed on p. 791. The Joule effect is irreversible, that is, heat is obtained in both directions of the current. In the Peltier effect, however, the effect is reversed when the current is reversed; that is, a cooling is produced at the junction of two metals in one direction, and an evolution of heat in the other direction.

Sir William Thomson, later Lord Kelvin, also found that heat was evolved or absorbed when a current flows along a metal whose ends are

kept at different temperatures. The *Thomson effect*, like the Peltier effect, is also reversible.

EXERCISES 32

1. State the laws of the development of heat when an electric current flows (a) through a wire of uniform material, (b) across the junction between two metals.

An electric heating coil is connected in series with a resistance of X ohms across the 240-volt mains, the coil being immersed in a kilogramme of water at 20°C . The temperature of the water rises to boiling-point in 10 minutes. When a second heating experiment is made with the resistance X short-circuited, the time required to develop the same quantity of heat is reduced to 6 minutes. Calculate the value of X . (Heat losses may be neglected.) (L.)

2. Define *electromotive force* and explain with the help of an example the difference between electromotive force and potential difference.

A thermocouple whose junctions are maintained at constant temperatures has a resistance of 5 ohms and its e.m.f. as measured using a potentiometer is 3.9 mV. What will be the reading on a millivoltmeter of resistance 60 ohms connected directly to the thermocouple? (N.)

3. Deduce an expression for the heat developed in a wire by the passage of an electric current.

The temperature of 300 g of paraffin oil in a vacuum flask rises 1.0°C per minute with an immersion heater of 12.3 watts input. On repeating with 400 g of oil the temperature rises by 1.20°C per minute for an input of 19.2 watts. Find the specific heat of the oil and the thermal capacity (assumed constant) of the flask. (L.)

4. Describe an experiment to determine the resistance of a wire by a calorimetric method.

It is desired to construct a 5-amp fuse from tin wire which has a melting-point of 230°C and resistivity 22×10^{-8} ohm m at that temperature. Estimate the diameter of the wire required if the emissivity of its surface is 88×10^{-5} J per sq. cm per second per $^{\circ}\text{C}$ excess temperature above the surroundings whose temperature is 20°C . Neglect the heat loss by conduction along the wire. (N.)

5. Describe the chief thermo-electric effects which occur in a circuit which includes two metals such as copper and iron.

Make a labelled diagram showing clearly the arrangement of a potentiometer circuit suitable for measuring a thermo-electric e.m.f. of about 2 mV. (L.)

6. Indicate, by means of graphs, the relation between the current and voltage (a) for a uniform manganin wire; (b) for a water voltameter; (c) for a diode valve. How do you account for the differences between the three curves?

An electric hot plate has two coils of manganin wire, each 20 metres in length and 0.23 mm^2 cross-sectional area. Show that it will be possible to arrange for three different rates of heating, and calculate the wattage in each case when the heater is supplied from 200-volt mains. The resistivity of manganin is 4.6×10^{-7} ohm m. (O. & C.)

7. Describe an experiment for determining the variation of the resistance of a coil of wire with temperature.

An electric fire dissipates 1 kW when connected to a 250-volt supply. Calculate to the nearest whole number the percentage change that must be made in the resistance of the heating element in order that it may dissipate 1 kW on a 200-volt supply. What percentage change in the length of the heating element will produce this change of resistance if the consequent increase in the temperature

of the wire causes its resistivity to increase by a factor 1.05? The cross-sectional area may be assumed constant. (*N.*)

8. What is a thermocouple? Explain the use of a potentiometer to measure the small electromotive forces developed by a thermocouple.

What are the relative advantages and disadvantages of a thermocouple used as a thermometer as compared with the resistance thermometer? (*L.*)

9. Give a general account of the thermo-electric effect. Describe how you would calibrate a thermocouple for use over the range of 0° – 100°C on the mercury-in-glass scale of temperature. (*C.*)

10. Derive, from first principles, an expression for the rate at which heat is generated in a resistance R by the passage of a current I .

An electric lamp takes 60 watts on a 240-volt circuit. How many dry cells, each of e.m.f. 1.45 volts and internal resistance 1.0 ohm, would be required to light the lamp? How much zinc would be consumed by the battery in 1 hour? (1 faraday = 96,500 coulombs; equivalent weight of zinc = 32.5.) (*O. & C.*)

11. Describe an experimental method of producing a thermo-electric e.m.f. How may a thermojunction be used to measure temperatures?

Why is a copper-iron junction not used to measure temperatures above 250°C , although a copper-constantan junction is often so employed? (*L.*)

12. Describe an instrument which measures the strength of an electric current by making use of its heating effect. State the advantages and disadvantages of this method.

A surge suppressor is made of a material whose conducting properties are such that the current passing through is directly proportional to the fourth power of the applied voltage. If the suppressor dissipates energy at a rate of 6.0 watts when the potential difference across it is 240 volts, estimate the power dissipated when the potential difference rises to 1200 volts. (*C.*)

13. Give a short account of the thermoelectric effect first discovered by Seebeck.

If you were given ice, boiling water, a thermocouple, a variable resistor and a sensitive galvanometer (with a linear scale) but no thermometer, describe how you would determine the temperature inside a domestic refrigerator. How would you test the assumption you are making? No other apparatus is available, but you may assume body temperature is 36.9°C .

Describe how you would measure, in the laboratory, the resistance of a pair of headphones, which is damaged if the current through it exceeds 100 milli-amp. (*C.*)

chapter thirty-three

Applications of Ohm's Law

MEASUREMENTS. NETWORKS

IN this chapter we shall apply Ohm's law to circuits more complicated than those of Chapter 32. We shall see that some special types of circuits can be used to make electrical measurements more accurately than with pointer instruments.

RESISTORS AND THEIR ARRANGEMENTS

Series Resistors

The resistors of an electric circuit may be arranged in series, so that the charges carrying the current flow through each in turn (Fig. 33.1); or they may be arranged in parallel, so that the flow of charge divides between them (Fig. 33.2), p. 808.

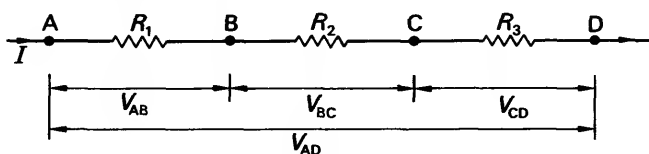


FIG. 33.1. Resistances in series.

Fig. 33.1 shows three passive resistors in series, carrying a current I . If V_{AD} is the potential difference across the whole system, the electrical energy supplied to the system per second is IV_{AD} . This is equal to the electrical energy dissipated per second in all the resistors; therefore

$$IV_{AD} = IV_{AB} + IV_{BC} + IV_{CD},$$

whence $V_{AD} = V_{AB} + V_{BC} + V_{CD},$ (1)

The individual potential differences are given by Ohm's law:

$$\left. \begin{aligned} V_{AB} &= IR_1 \\ V_{BC} &= IR_2 \\ V_{CD} &= IR_3 \end{aligned} \right\} (2)$$

and

Hence, by equation (1),

$$\begin{aligned} V_{AD} &= IR_1 + IR_2 + IR_3 \\ &= I(R_1 + R_2 + R_3). \end{aligned} (3)$$

And the effective resistance of the system is

$$R = \frac{V_{AD}}{I} = R_1 + R_2 + R_3. \quad (4)$$

The physical facts are:

- (i) *Current same through all resistors.*
- (ii) *Total potential difference = sum of individual potential difference (equation (1)).*
- (iii) *Individual potential differences directly proportional to individual resistances (equation (2)).*
- (iv) *Total resistance greater than greatest individual resistance (equation (4)).*
- (v) *Total resistance = sum of individual resistances.*

Resistors in Parallel

Fig. 33.2 shows three passive resistors connected in parallel, between the points A, B. A current I enters the system at A and leaves at B,

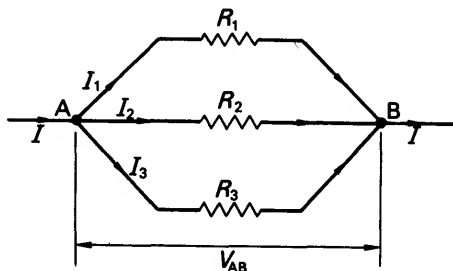


FIG. 33.2. Resistances in parallel.

setting up a potential difference V_{AB} between those points. The current branches into I_1, I_2, I_3 , through the three elements, and

$$I = I_1 + I_2 + I_3. \quad (5)$$

Now
$$I_1 = \frac{V_{AB}}{R_1}, \quad I_2 = \frac{V_{AB}}{R_2}, \quad I_3 = \frac{V_{AB}}{R_3}.$$

$$\therefore I = V_{AB} \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right).$$

$$\therefore \frac{I}{V_{AB}} = \frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}. \quad (6)$$

where R is the effective resistance (V_{AB}/I) of the system.

The physical facts about resistors in parallel may be summarized as follows:

- (i) *Potential difference same across each resistor.*
- (ii) *Total current = sum of individual currents (equation (5)).*

- (iii) *Individual currents inversely proportional to individual resistances.*
- (iv) *Effective resistance less than least individual resistance (equation (6)).*

Resistance Boxes

In many electrical measurements variable known resistances are required; they are called resistance boxes. As shown in Fig. 33.3 (i) ten coils, each of resistance 1 ohm, for example, are connected in series. A rotary switch with eleven contacts enables any number of these coils to be connected between the terminals AA'. A resistance box contains several sets of coils and switches, the first giving resistances 0–10 ohms in steps of 1 ohm, the next 0–100 ohms in steps of 10 ohms, and so on. These are called decade boxes.

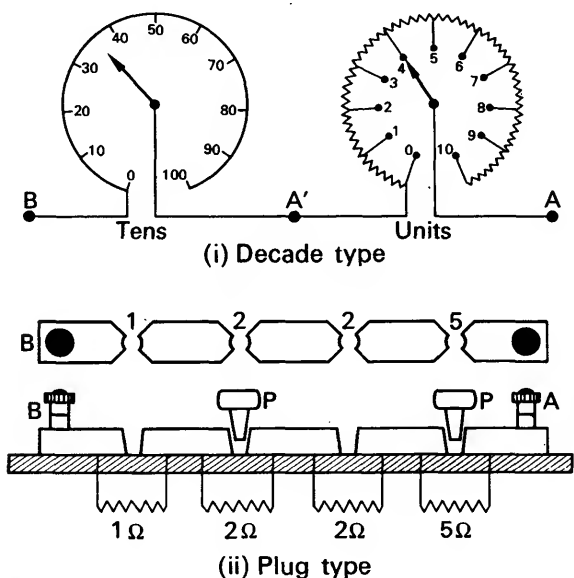


FIG. 33.3. Resistance boxes.

The switches used in a decade box are of very high quality; their contact resistances are negligible compared with the resistances of the coils which they select. Switches of this kind have been developed only in the last twenty years or so: in older boxes no switches are used. Instead, the resistances are varied by means of plugs. As shown in Fig. 33.3 (ii), the resistance coils are joined across gaps in a thick brass bar, and the gaps are formed into tapered sockets to receive short-circuiting plugs P. The resistance between the terminals A and B in Fig. 33.3 (ii) is the sum of the unplugged resistances between them—3 ohms in this example.

The coils of a resistance box are wound in a particular way, which we shall describe and explain later (p. 924). They are not intended to carry large currents, and must not be allowed to dissipate more than one watt. Therefore, since $P = I^2 R$, the greatest safe current for a 1-ohm coil is 1 amp, and for a 10-ohm coil about 0.3 amp. If the one-watt limit is exceeded, the insulation will be damaged, or the wire burnt out.

The Potential Divider

Two resistance boxes in series are often used in the laboratory to provide a known fraction of a given potential difference—for example, of one which is too large to measure easily. Fig. 33.4 (i) shows the

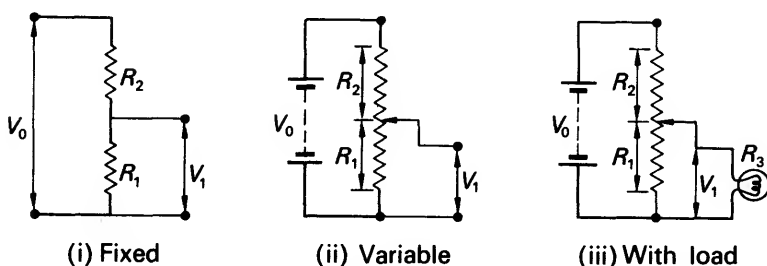


FIG. 33.4. Potential divider.

arrangement, which is called a resistance 'potential divider'. The current flowing, I , is given by

$$I = \frac{V_0}{R_1 + R_2},$$

$$\therefore V_1 = IR_1 = \frac{R_1}{R_1 + R_2} V_0. \quad (7)$$

A resistor with a sliding contact can similarly be used, as shown in Fig. 33.4(ii), to provide a continuously variable potential difference, from zero to the full supply value V_0 . This is a convenient way of controlling the voltage applied to a load, such as a lamp (Fig. 33.4(iii)). The resistance of the load, R_3 , however, acts in parallel with the resistance R_1 ; equation (7) is therefore no longer true, and the voltage V_1 must be measured with a voltmeter. It can be calculated, as in the following example, if R_3 is known; but if the load is a lamp its resistance varies greatly with the current through it, because its temperature varies.

EXAMPLE

A load of 2000 ohms is connected, via a potential divider of resistance 4000 ohms, to a 10-volt supply (Fig. 33.5). What is the potential difference across the load when the slider is (a) one-quarter, (b) half-way up the divider?

$$\text{Since } \frac{1}{R} = \frac{1}{2000} + \frac{1}{1000}$$

$$(a) R_{BC} = \frac{2000 \times 1000}{2000 + 1000} = \frac{2000}{3} \text{ ohms.}$$

$$\therefore R_{AC} = R_{AB} + R_{BC} = 3000 + \frac{2000}{3} \text{ ohms,}$$

$$\begin{aligned} \therefore V_{BC} &= \frac{R_{BC}}{R_{AC}} V_C \\ &= \frac{2000/3}{11000/3} \times 10 = \frac{2}{11} \times 10 \\ &= 1.8 \text{ volts.} \end{aligned}$$

If the load were removed, V_{BC} would be 2.5 volts.

(b) It is left for the reader to show similarly that $V_{BC} = 3.3$ volts. Without the load it would be 5 volts.

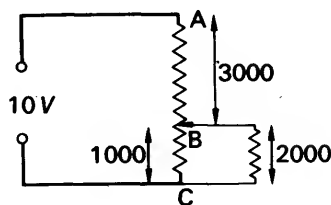


FIG. 33.5. A loaded potential divider.

MEASURING INSTRUMENTS

Conversion of a Millimeter into a Voltmeter

Ohm's law enables us to use a millimeter as a voltmeter. Let us suppose that we have a moving-coil instrument which requires 5 milliamperes for full-scale deflection (f.s.d.). And let us suppose that the resistance of its coil, r , is 20 ohms (Fig. 33.6). Then, when it is fully deflected, the potential difference across it is

$$\begin{aligned} V &= rI \\ &= 20 \times 5 \times 10^{-3} = 100 \times 10^{-3} \text{ volt} \\ &= 0.1 \text{ volt.} \end{aligned}$$

Since the coil obeys Ohm's law, the current through it is proportional to the potential difference across it; and since the deflection of the pointer is proportional to the current it is therefore also proportional to the potential difference. Thus the instrument can be used as a voltmeter, giving full-scale deflection for a potential difference of 0.1 volt, or 100 millivolts. Its scale could be engraved as shown at the top of Fig. 33.6.

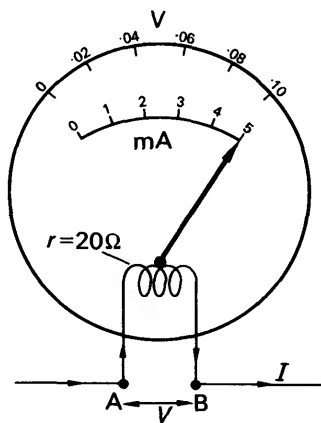
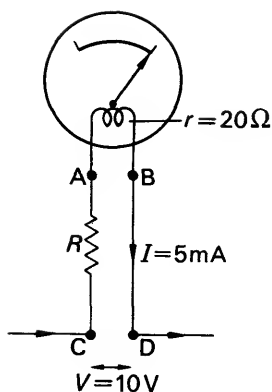


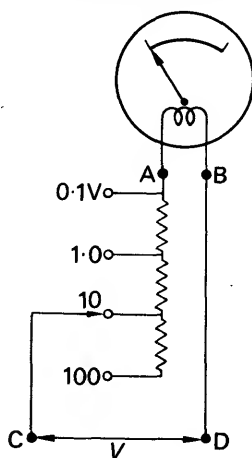
FIG. 33.6. P.D. across moving-coil meter.

The potential differences to be measured in the laboratory are usually greater than 100 millivolts, however. To measure such a potential difference, we insert a resistor R in series with the coil, as shown in Fig. 33.7. If we wish to measure up to 10 volts we must choose the resistance R so that, when 10 volts are applied between the



Single-range

FIG. 33.7. Single-range voltmeter.



Multi-range

FIG. 33.8. Multi-range voltmeter.

terminals CD, then a current of 5 milliamperes flows through the moving coil. By Ohm's law

$$V = (R + r)I,$$

$$\therefore 10 = (R + 20) \times 5 \times 10^{-3}$$

$$\text{or} \quad R + 20 = \frac{10}{5 \times 10^{-3}} = 2 \times 10^3 = 2000 \text{ ohms.}$$

$$\therefore R = 2000 - 20$$

$$= 1980 \text{ ohms.} \quad \dots \dots \dots (8)$$

The resistance R is called a *multiplier*. Many voltmeters contain a series of multipliers of different resistances, which can be chosen by a switch or plug-and-socket arrangement (Fig. 33.8).

Conversion of a Milliammeter into an Ammeter

Moving-coil instruments give full-scale deflection for currents smaller than those generally encountered in the laboratory. If we wish to measure a current of the order of an ampere or more we connect a low resistance S , called a *shunt*, across the terminals of a moving-coil meter (Fig. 33.9). The shunt diverts most of the current to be measured, I , away from the coil—hence its name. Let us suppose that, as before,

the coil of the meter has a resistance r of 20 ohms and is fully deflected by a current, I_C of 5 milliamperes. And let us suppose that we wish to shunt it so that it gives f.s.d. for 5 amperes to be measured. Then the current through the shunt is

$$\begin{aligned} I_s &= I - I_C \\ &= 5 - 0.005 \\ &= 4.995 \text{ amp.} \end{aligned}$$

The potential difference across the shunt is the same as that across the coil, which is

$$V = rI_C = 20 \times 0.005 = 0.1 \text{ volt.}$$

The resistance of the shunt must therefore be

$$S = \frac{V}{I_s} = \frac{0.1}{4.995} = 0.02002 \text{ ohms.} \quad (9)$$

The ratio of the current measured to the current through the coil is

$$\frac{I}{I_C} = \frac{5}{5 \times 10^{-3}} = 1000.$$

This ratio is the same whatever the current I , because it depends only on the resistances S and r ; the reader may easily show that its value is $(S+r)/S$. The deflection of the coil is therefore proportional to the measured current, as indicated in the figure, and the shunt is said to have a 'power' of 1000 when used with this instrument.

The resistance of shunts and multipliers are always given with four-figure accuracy. The moving-coil instrument itself has an error of the order of 1 per cent.; a similar error in the shunt or multiplier would therefore double the error in the instrument as a whole. On the other hand, there is nothing to be gained by making the error in the shunt less than about 0.1 per cent., because at that value it is swamped by the error of the moving system.

Multimeters

A *multimeter* instrument is one which is adapted for measuring both current and voltage. It has a shunt R as shown, and a series of voltage multipliers R' (Fig. 33.10). The shunt is connected permanently across the coil, and the resistances in R' are adjusted to give the desired full-scale voltages with the shunt in position. A switch or plug enables the various full-scale values of current or voltage to be chosen, but the user does the mental arithmetic. The instrument shown in the figure is reading 1.7 volts; if it were on the 10-volt range, it would be reading 6.4.

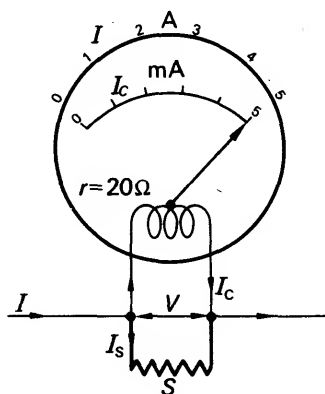


FIG. 33.9. Conversion of milliammeter to ammeter.

The terminals of a meter, multimeter or otherwise, are usually marked + and -; the pointer is deflected to the right when current passes through the meter from + to -.

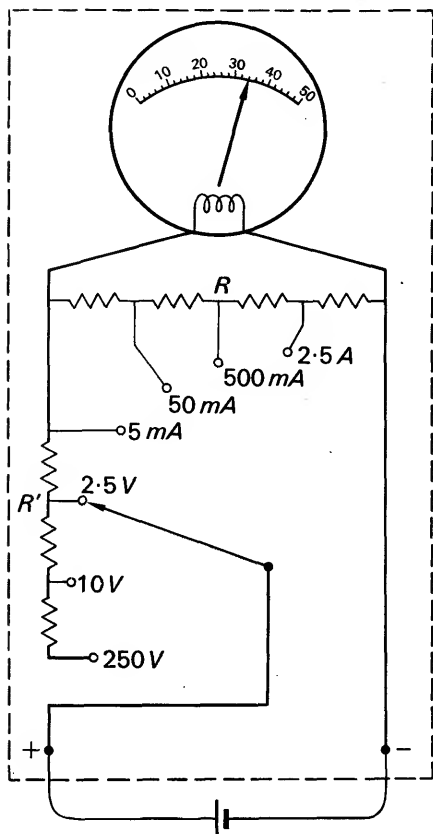


FIG. 33.10. A multimeter.

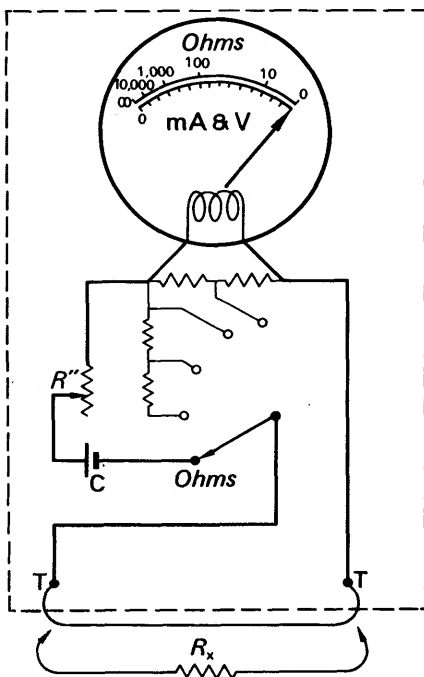


FIG. 33.11. Measurement of resistance with multimeter.

The multimeters are generally arranged to measure resistance as well as current and voltage. An extra position on the switch, marked 'R' or 'ohms', puts a dry cell C and a variable resistor R'' in series with the moving coil (Fig. 33.11). Before the instrument is used to measure a resistance, its terminals TT are short-circuited, and R'' is adjusted until the pointer is fully deflected. As shown in the figure, it is then opposite the zero on the ohms scale. The short-circuit is next removed, and the unknown resistance R_x is connected across the terminals. The current falls, and the pointer moves to the left, indicating on the ohms scale the value of R_x . The ohms scale is calibrated by the makers with known resistances.

Use of Voltmeter and Ammeter

A moving-coil voltmeter is a current-operated instrument. It can be used to measure potential differences only because the current which it

draws is proportional to the potential difference applied to it, from Ohm's law. Since its action depends on Ohm's law, a moving-coil voltmeter cannot be used in any experiment to demonstrate that law; that is why, when describing such an experiment on p. 789, we specified measuring instruments whose readings are not dependent on Ohm's law.

Having once established Ohm's law, however, we can use moving-coil voltmeters freely; they are both more sensitive and more accurate than other forms of voltmeters. The current which they take does, however, sometimes complicate their use. To see how it may do so, let us suppose that we wish to measure a resistance R of about 100 ohms. As shown in Fig. 33.12, we connect it in series with a cell, a milliammeter, and a

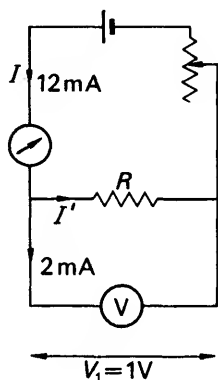


FIG. 33.12

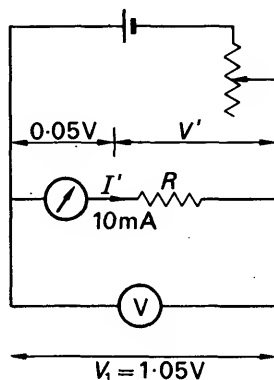


FIG. 33.13.

Use of ammeter and voltmeter.

variable resistance; across it we place the voltmeter. We adjust the current until the voltmeter reads, say, $V_1 = 1$ volt; let us suppose that the milliammeter then reads $I = 12$ mA. The value of the resistance then appears to be

$$R = \frac{V_1}{I} = \frac{1}{12 \times 10^{-3}} = \frac{10^3}{12} \\ = 83 \text{ ohms (approx.).}$$

But the milliammeter reading includes the current drawn by the voltmeter. If that is 2 mA, then the current through the resistor, I' , is only 10 mA and its resistance is actually

$$R = \frac{V_1}{I'} = \frac{1}{10 \times 10^{-3}} = \frac{1}{10^{-2}} \\ = 100 \text{ ohms.}$$

The current drawn by the voltmeter has made the resistance appear 17 per cent. lower than its true value.

In an attempt to avoid this error, we might connect the voltmeter as shown in Fig. 33.13: across both the resistor and the milliammeter. But

its reading would then include the potential difference across the milliammeter. Let us suppose that this is 0.05 volt when the current through the milliammeter is 10 mA. Then the potential difference V' across the resistor would be 1 volt, and the voltmeter would read 1.05 volt. The resistance would appear to be

$$R = \frac{1.05}{10 \times 10^{-3}} = \frac{1.05}{10^{-2}} \\ = 105 \text{ ohms.}$$

Thus the voltage drop across the milliammeter would make the resistance appear 5 per cent. higher than its true value.

Errors of this kind are negligible only when the voltmeter current is much less than the current through the resistor, or when the voltage drop across the ammeter is much less than the potential difference across the resistor. If we were measuring a resistance of about 1 ohm, for example, the current I in Fig. 33.12 would be 1 amp, and I would be 1.002 amp. The error in measuring R would then be only 0.2 per cent—less than the intrinsic error of the meter. But the circuit of Fig. 33.13 would give the same error as before. It could do so because, as we saw when considering shunts, the shunt across the milliammeter would have been chosen to make the voltage drop still 0.05 volt. Thus V_1 would still be 1.05 volt when V' was 1 volt, and the error would be 5 per cent as before.

In low-resistance circuits, therefore, the voltmeter should be connected as in Fig. 33.12, so that its reading does not include the voltage drop across the ammeter. But in high-resistance circuits the voltmeter should be connected as in Fig. 33.13, so that the ammeter does not carry its current.

If a moving-coil voltmeter is connected across a cell, it will not read its true e.m.f., because the current which it draws will set up a voltage drop across the internal resistance of the cell. The drop will be negligible only if the resistance of the voltmeter is very high compared with the internal resistance. E.m.f.s are thus compared by a potentiometer method, discussed shortly.

Figure of Merit of a Voltmeter

If a milliammeter of 1 mA f.s.d. (full scale deflection) is converted into a voltmeter, then if it is to have 1 volt f.s.d. its total resistance—coil plus multiplier—must be 1000 ohms. (One volt across its terminals will send through it a current of $1/1000$ amp = 1 mA.) If it is to have 10 volts f.s.d., then its total resistance must be 10000 ohms; for 20 volts f.s.d., 20000 ohms, and so on. It will have a resistance of 1000 ohms for every volt of its full-scale deflection. Such a meter is said to have a *figure of merit* of 1000 ohms per volt. Similarly, a voltmeter which takes 10 mA, or $1/100$ amp, for full-scale deflection has a figure of merit of 100 ohms per volt. The greater the figure of merit of a voltmeter, expressed in this way, the less will it disturb any circuit to which

it is connected, and the less error will its current cause in any measurements made with it. On the other hand, the greater the figure of merit, the more delicate the moving system of the meter and the greater its intrinsic error. First-grade, and particularly 'sub-standard', meters therefore have medium or low figures of merit: from 500 to 66.7 ohms per volt.

When a voltmeter of low figure of merit is being used, it may be necessary to allow for the current which it draws. The allowance is made in the way indicated on p. 815, where the use of a voltmeter and ammeter together was discussed.

THE POTENTIOMETER

Pointer instruments are useless for very accurate measurements: the best of them have an intrinsic error of about 1 per cent of full scale. Where greater accuracy than this is required, elaborate measuring circuits are used.

One of the most versatile of these, due to Poggendorf, is the potentiometer. It consists of a uniform wire, AB in Fig. 33.13(i), about a metre long; through it an accumulator X maintains a steady current I . Since the wire is uniform, its resistance per centimetre, R , is constant; the voltage drop across 1 cm of the wire, RI , is therefore also constant.

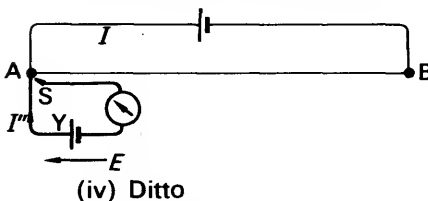
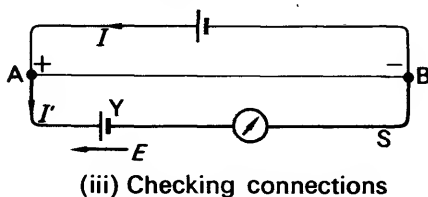
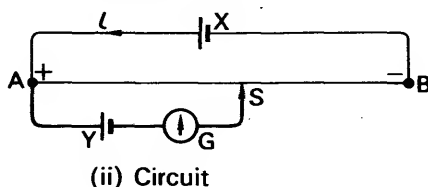
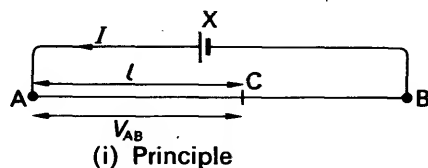


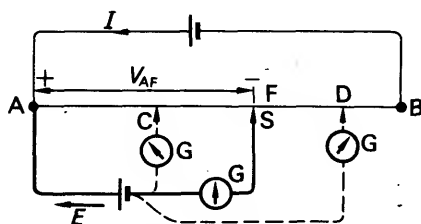
FIG. 33.14. The potentiometer.

The potential difference between the end A of the wire, and any point C upon it, is thus proportional to the length of wire l between A and C:

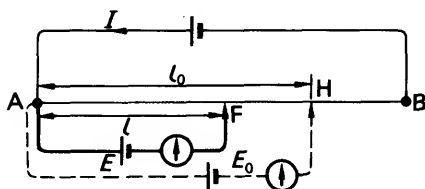
$$V_{AC} \propto l. \quad (10)$$

Comparison of E.M.F.s

To illustrate the use of the potentiometer, let us suppose that we take a cell, Y in Fig. 33.14(ii), and join its positive terminal to the point A (to which the positive terminal of X is also joined). We connect the negative terminal of Y, via a sensitive galvanometer, to a slider S, which we can press on to any point in the wire. Let us suppose that the cell Y has an e.m.f. E , which is less than the potential difference V_{AB} across the whole of the wire. Then if we press the slider on B, a current I' will flow through Y in opposition to its e.m.f. (Fig. 33.14(iii)). This current will deflect the galvanometer G—let us say to the right. If we now press the slider on A, the cell Y will be connected straight across the galvanometer, and will deliver a current I'' in the direction of its e.m.f. (Fig. 33.14(iv)). The galvanometer will therefore show a deflection to the left. If the deflections at A and B are not opposite, then either the e.m.f. of Y is greater than the potential difference across the whole wire, or we have connected the circuit wrongly. The commonest mistake in connecting up is not joining both positives to A.



(i) Finding balance point



(ii) Comparison of e.m.f.

FIG. 33.15. Use of potentiometer.

Now let us suppose that we place the slider on to the wire at a point a few centimetres from A, then at a point a few centimetres farther on, and so forth. (We do not run the slider continuously along the wire, because the scraping would destroy the uniformity.) When the slider is at a point C near A (Fig. 33.15(i)) the potential difference V_{AC} is less than the e.m.f. E of Y; current therefore flows through G in the direction

of E , and G may deflect to the left. When the slider is at D near B , V_{AD} is greater than E , current flows through G in opposition to E , and G deflects to the right. By trial and error (but no scraping of the slider) we can find a point F such that, when the slider is pressed upon it, the galvanometer shows no deflection. The potential difference V_{AF} is then equal to the e.m.f. E ; no current flows through the galvanometer because E and V_{AF} act in opposite directions in the galvanometer circuit (Fig. 33.15(i)). Because no current flows, the resistance of the galvanometer, and the internal resistance of the cell, cause no voltage drop; the full e.m.f. E therefore appears, between the points, A and S , and is balanced by V_{AF} :

$$E = V_{AF}.$$

If we now take another cell of e.m.f. E_0 , and balance it in the same way, at a point H (Fig. 33.15(ii)), then

$$E_0 = V_{AH}.$$

Therefore
$$\frac{E}{E_0} = \frac{V_{AF}}{V_{AH}}.$$

The potential differences V_{AF} , V_{AH} are proportional to the lengths l , l_0 from A to F , and from A to H , respectively. Therefore

$$\frac{E}{E_0} = \frac{l}{l_0}. \quad (11)$$

Accuracy

When the potentiometer is used to compare the e.m.f.s of cells, no errors are introduced by the internal resistances, because no current flows at the balance-points.

The potentiometer is more accurate than an electrometer instrument, which, like a moving-coil voltmeter, has an intrinsic error of about 1 per cent of full-scale. The accuracy of a potentiometer is limited by the non-uniformity of the slide-wire, the uncertainty of the balance-point, and the error in measuring the length l of wire from the balance-point to the end A . With even crude apparatus, the balance-point can be located to within about 0.5 mm; if the length l is 50 cm, or 500 mm, then the error in locating the balance-point is 1:1000. If the wire has been carefully treated, its non-uniformity may introduce an error of about the same magnitude. The overall error is then about ten times less than that of a pointer instrument. A refined potentiometer has a still smaller error.

The precision with which the balance-point of a potentiometer can be found depends on the sensitivity of the galvanometer—the smallness of the current which will give a just-discernible deflection. A moving-coil galvanometer must be protected by a series resistance R of several thousand ohms, which is shorted out when the balance is nearly reached (Fig. 33.16). A series resistance is preferable to a shunt, because it re-

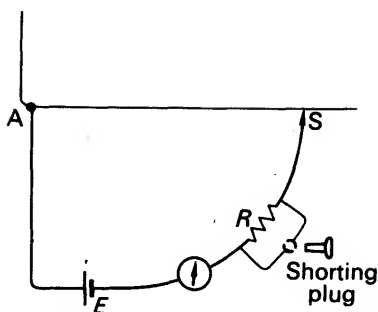


FIG. 33.16. Use of protective resistance with galvanometer.

duces the current drawn from the cell under test, when the potentiometer is unbalanced. The process of seeking the balance-point then causes less change in the chemical condition of the cell, and therefore in its e.m.f.

It is important to realize that the accuracy of a potentiometer does not depend on the accuracy of the galvanometer, but only on its sensitivity. The galvanometer is used not to measure a

current but merely to show one when the potentiometer is off balance. It is said to be used as a null-indicator, and the potentiometer method of measurement, like the bridge methods which we shall describe shortly, is called a null method.

The current through the potentiometer wire must be steady—it must not change appreciably between the finding of one balance-point and the next. The accumulator which provides it should therefore be neither freshly charged nor nearly run-down; when an accumulator is in either of those conditions its e.m.f. falls with time. Errors in potentiometer measurements may be caused by non-uniformity of the wire, and by the resistance of its connexion to the terminal at A. This resistance is added to the resistance of the length l of the wire between A and the balance-point, and if it is appreciable it makes equation (11) invalid. Both these sources of error are eliminated in the Rayleigh potentiometer, which we shall describe later (p. 825).

Uses of the Potentiometer. E.M.F. and Internal Resistance

All the uses of the potentiometer depend on the fact that it can measure potential difference accurately, and without drawing current from the circuit under test.

If one of the cells in Fig. 33.15 (ii) has a known e.m.f., say E_0 then the e.m.f. of the other, E , is given by equation:

$$\frac{E}{E_0} = \frac{l}{l_0} \quad (12)$$

A cell of known e.m.f. is called a standard cell. The e.m.f.s of standard cells are determined absolutely—that is to say, without reference to the e.m.f.s of any other cell—by methods which depend, in principle, on the definition of e.m.f. (power/current, p. 797). Standard cells are described on p. 865, along with the precautions which must be taken in their use. For simple experiments a Daniell cell (p. 860), whose e.m.f. is about 1.1 volt, may be used as a standard.

Equation (12) is true only if the current I through the potentiometer wire has remained constant. The easiest way to check that it has done so is to balance the standard cell against the wire before and after balancing the unknown cell. If the lengths to the balance-point are

equal—within the limits of experimental error— then the current I may be taken as constant. A check of this kind should be made in each of the experiments to be described.

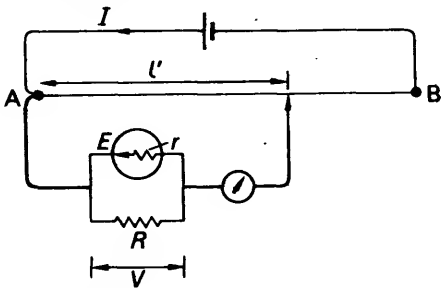


FIG. 33.17. Measurement of internal resistance.

The internal resistance of a cell, r , can be found with a potentiometer by balancing first its e.m.f., E , and the its terminal potential difference, V , when a known resistance R is connected across it (Fig. 33.17). Ohm's law for the complete circuit gives

$$\frac{V}{E} = \frac{R}{R + r} \quad \dots \quad (13)$$

But

$$\frac{V}{E} = \frac{l}{l'} \quad \dots \quad (14)$$

where l and l' are the lengths of potentiometer wire required to balance E and V . From equations (13) and (14), r can be found from

$$r = \left(\frac{l}{l'} - 1 \right) R.$$

Calibration of Voltmeter

Fig. 33.18 shows how a potentiometer can be used to calibrate a voltmeter. A standard cell is first used to find the p.d. per cm or volts

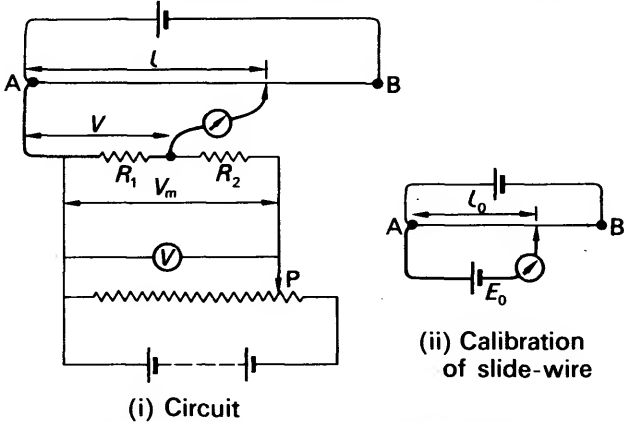


FIG. 33.18. Calibration of voltmeter with potentiometer.

per cm of the wire (Fig. 33.18(ii)): if its e.m.f. E_0 is balanced by a length l_0 , then

$$\text{volts per cm} = \frac{E_0}{l_0}. \quad (15)$$

Different voltages V_m are now applied to the voltmeter by the adjustable potential divider P (Fig. 33.18(i)). The fixed potential divider, comprising $R_1 R_2$, gives a known fraction V of each value of V_m , which is then balanced on the potentiometer:

$$\frac{V}{V_m} = \frac{R_1}{R_1 + R_2}. \quad (16)$$

(The resistances R_1 and R_2 are high—of the order of 1000 to 10000 ohms, so that the voltage adjustment by P is fairly uniform. Their ratio is chosen so that the greatest value of V is measurable on the potentiometer—about 1.5 volts.) If l is the lengths of potentiometer wire which balances a given value of V , then

$$\begin{aligned} V &= l \times (\text{volt/cm of wire}) \\ &= l \frac{E_0}{l_0}. \end{aligned}$$

From each value of V , the value of V_m is calculated by equation (16). If the voltmeter reading is V_{obs} , then the correction to be added to it is $V_m - V_{obs}$. This is plotted against V_{obs} as in Fig. 33.19.

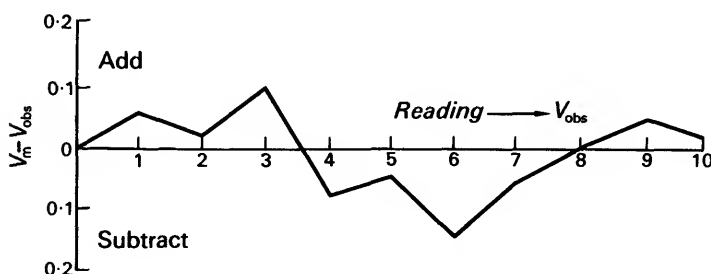


FIG. 33.19. Correction curve of voltmeter.

Measurement of Current

A current can be measured on a potentiometer by means of the potential difference which it sets up across a known resistance, R in Fig. 33.20(i). The resistance is low, being chosen so that the potential difference across it is of the order of 0.1 volt. (A higher value is not chosen, because the voltage drop across the resistor disturbs the circuit in which it is inserted.) Fig. 33.20(ii) shows in detail the kind of resistor used, which is often called a standard shunt. It consists of a broad strip of alloy, such as manganin, whose resistance varies very little with temperature (p. 837). The current is led in and out as the terminals i, i . The terminals v, v are connected to fine wires soldered to points PP

on the strip; they are called the potential terminals. The marked value R of the resistance is the value between the points PP; it is adjusted by making hack-saw cuts into the edges of the strip.

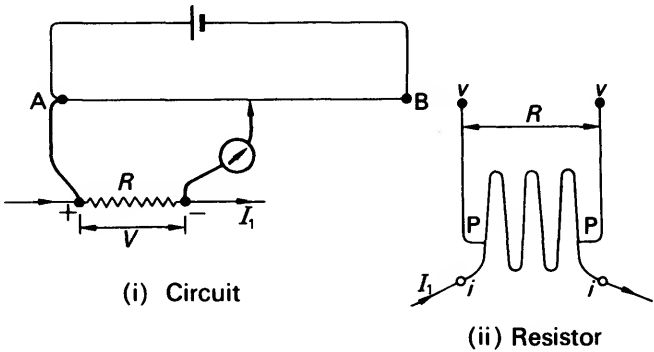


FIG. 33.20. Measurement of current with potentiometer.

As shown in Fig. 33.20(i), the current to be measured, I_1 , is passed through the shunt, and the potential difference between its potential terminals, V , is balanced on the potentiometer wire. If l is the length of wire to the balance-point, then

$$\frac{V}{E_0} = \frac{l}{l_0}, \quad \dots \dots \dots (17)$$

where E_0 and l_0 refer to a standard cell as before. Equation (17) enables the current to be found in terms of E_0 , l_0 , l , and R , since

$$V = I_1 R.$$

The resistance of the wires connecting the potential terminals to the points PP, and to the potentiometer circuit, do not affect the result, because at the balance-point the current through them is zero.

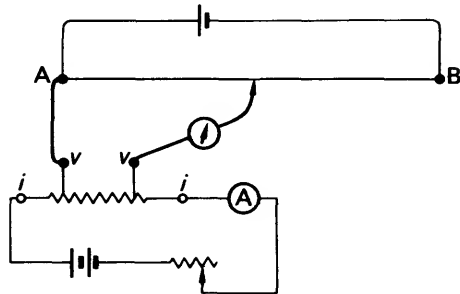


FIG. 33.21. Calibration of ammeter with potentiometer.

This method of measuring a current can be used to calibrate an ammeter, A. The circuit is shown in Fig. 33.21; its principle and use

should explain themselves. The results are treated in the same way as in the calibration of a voltmeter.

Comparison of Resistances

A potentiometer can be used to compare resistances, by comparing the potential differences across them when they are carrying the same current I_1 (Fig. 33.22). This method is particularly useful for very low resistances, because, as we have just seen, the resistances of the connecting wires do not affect the result of the experiment. It can, however, be used for higher resistances if desired. With low resistances the ammeter A' and rheostat P are necessary to adjust the current to a value which will neither exhaust the accumulator Y , nor overheat the resistors. No standard cell is required. The potential difference across the first resistor, $V_1 = R_1 I_1$, is balanced against a length l_1 of the potentiometer wire, as shown by the full lines in the figure. Both potential terminals

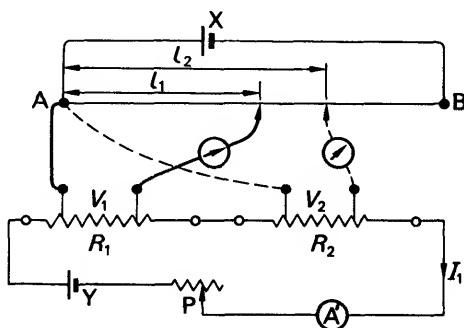


FIG. 33.22.
Comparison of resistances
with potentiometer.

of R_1 are then disconnected from the potentiometer, and those of R_2 are connected in their place. If l_2 is the length to the new balance-point, then

$$\frac{l_1}{l_2} = \frac{V_1}{V_2} = \frac{R_1 I_1}{R_2 I_1} = \frac{R_1}{R_2}.$$

This result is true only if the current I_1 is constant; as well as the potentiometer current. The accumulator Y , as well as X , must therefore be in good condition. To check the constancy of the current I_1 , the ammeter A' is not accurate enough. The reliability of the experiment as a whole can be checked by balancing the potential V_1 a second time, after V_2 . If the new value of l_1 differs from the original, then at least one of the accumulators is running down and must be replaced.

The Rayleigh Potentiometer

Fig. 33.23 shows a potentiometer devised by Lord Rayleigh (1842–1919), which is free from errors due to non-uniformity of the wire and to contact resistance at the end A (p. 820). It consists of two plug-type resistance boxes, R_1 , R_2 , joined in series. (These boxes may well be the R -sections of two similar Post Office boxes (p. 834). At the start of a measurement all the plugs of R_1 are inserted, and all of R_2 taken out. Then R_1 is zero, and the main current I sets up no potential difference across it; but when the key K is pressed, the unknown e.m.f. E deflects the galvanometer. R_1 is now increased by, say, 100 ohms, and R_2 is

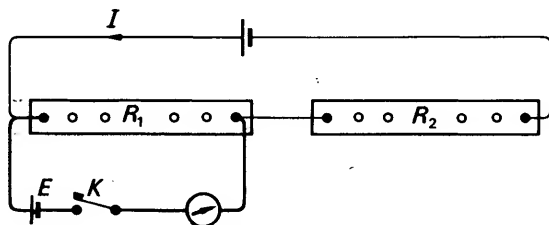


FIG. 33.23. Rayleigh potentiometer.

decreased by the same amount. In this way $R_1 + R_2$ is kept constant, and the current I does not change. But there is now a potential difference across R_1 , which opposes E . Plugs are taken out of R_1 and put into R_2 , so as to keep $R_1 + R_2$ constant, until the galvanometer shows no deflection when K is pressed. If R'_1 is the value of R_1 at this point, then

$$E = R'_1 I.$$

The procedure is now repeated with a standard cell of e.m.f. E_0 , in place of E . Since $R_1 + R_2$ has been kept constant, the current I is the same as before; hence, if R''_1 is the new value of R_1 at balance,

$$E_0 = R''_1 I.$$

Consequently,

$$\frac{E}{E_0} = \frac{R'_1 I}{R''_1 I} = \frac{R'_1}{R''_1}.$$

Measurement of Thermal E.M.F.

The e.m.f.s of thermojunctions (p. 803) are small—of the order of a millivolt. If we attempted to measure such an e.m.f. on a simple potentiometer we should find the balance-point very near one end of the wire, so that the end-error would be serious. The Rayleigh potentiometer, although it is free from end-errors, is not suitable for measuring small e.m.f.s; if, in Fig. 33.23, $R_1 + R_2 = 10000 \, \Omega$, and the e.m.f.

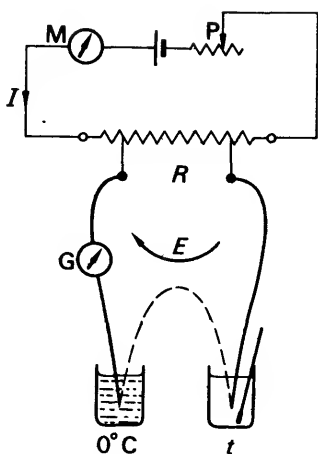


FIG. 33.24. Measurement of thermal e.m.f.

of the accumulator = 2 volts, then $I = 2/10000 = 2 \times 10^{-4}$ amp = 0.2 mA. To balance a thermal e.m.f. of 2 mV, R_1 would therefore have to be 10Ω ; and since R_1 cannot be adjusted in steps smaller than 1Ω , the e.m.f. cannot be measured to a greater accuracy than 10 per cent.

For accurate measurement of thermal e.m.f.s special potentiometers have been devised, but the simple circuit of Fig. 33.24 will do for a laboratory experiment. The e.m.f. E is applied via a sensitive galvanometer G across a standard shunt R of about 1 ohm. A current I , of a few milliamperes, is passed through the shunt, and measured on the millimeter M . Its value is adjusted by the rheostat P until G

shows no deflection. The potential difference RI is then equal and opposite to the thermal e.m.f.

$$E = RI.$$

If a balance cannot be found, the connexions of the junction to R should be reversed.

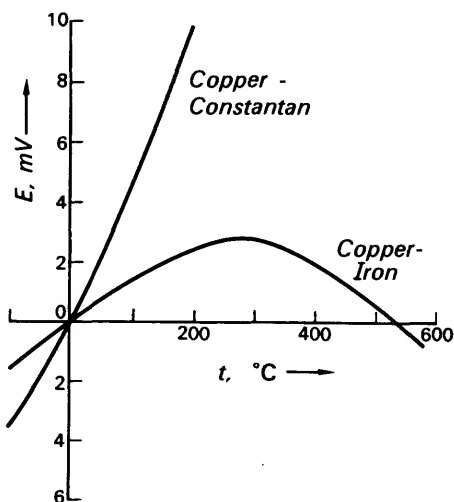


FIG. 33.25. E.m.f.s of thermocouples.
(Reckoned positive when into copper at the cold junction.)

Fig. 33.25 shows the results of measuring the e.m.f. E when the cold junction is at 0°C and the hot is at various temperatures t . The curves approximate to parabolas:

$$E = at + bt^2. \quad (18)$$

THERMO-ELECTRIC E.M.F.s

(*E* in micro-volts when *t* is in °C
and cold junction at 0°C)

Junction	<i>a</i>	<i>b</i>	Range for <i>a</i> and <i>b</i> , °C	Limits of use, °C
Cu/Fe	14	−0.02	0–100	See 1
Cu/Constantan ²	41	0.04	−50 to +300	−200 to +300
Pt/Pt–Rh ³	6.4	0.006	0–200	0–1700
Chromel ⁴ /Alumel ⁵	41	0.00	0–900	0–1300

¹ Simple demonstrations.
² See p. 788.
³ 10 per cent Rh; used only for accurate work or very high temperatures.
⁴ 90 per cent Ni, 10 per cent Cr.
⁵ 94 per cent Ni, 3 per cent Mn, 2 per cent Al, 1 per cent Si.

NETWORKS

Kirchhoff’s Laws

A ‘network’ is usually a complicated system of electrical conductors. Kirchhoff (1824–87) extended Ohm’s law to networks, and gave two laws, which together enabled the current in any part of the network to be calculated.

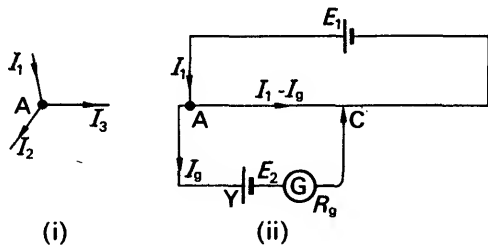


FIG. 33.26. Kirchhoff’s laws.

The *first law* refers to any point in the network, such as A in Fig. 33.26 (a); it states that the total current flowing into the point is equal to the total current flowing out of it:

$$I_1 = I_2 + I_3.$$

The law follows from the fact that electric charges do not accumulate at the points of a network. It is often put in the form that *the algebraic sum of the currents at a junction of a circuit is zero*, or

$$\Sigma I = 0,$$

where a current, *I*, is reckoned positive if it flows towards the point, and negative if it flows away from it. Thus at A in Fig. 33.26 (i),

$$I_1 - I_2 - I_3 = 0.$$

Kirchhoff's first law gives a set of equations which contribute towards the solving of the network; in practice, however, we can shorten the work by putting the first law straight into the diagram, as shown in Fig. 33.26 (ii) for example, since

$$\text{current along AC} = I_1 - I_g.$$

Kirchhoff's *second law* is a generalization of Ohm's law for the complete circuit. It refers to any closed loop, such as AYCA in Fig. 33.26 (ii); and it states that, round such a loop, the algebraic sum of the voltage drops is equal to the algebraic sum of the e.m.f.s:

$$\Sigma RI = \Sigma E.$$

Thus, clockwise round the loop,

$$R_{AC}(I_1 - I_g) - R_g I_g = E_2.$$

We have used the potentiometer to illustrate Kirchhoff's laws merely because it is already familiar to us; we shall not go on and solve it as a network, because we have already dealt with as much of the theory of it as we need.

EXAMPLE

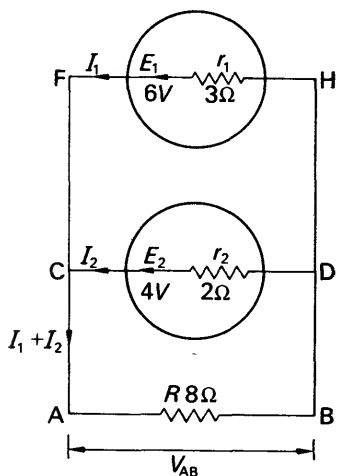


FIG. 33.27. Load across cells in parallel.

Fig. 33.27 shows a network which can be solved by Kirchhoff's laws. From the first law, the current in the 8-ohm wire is $(I_1 + I_2)$, assuming I_1 , I_2 are the currents through the cells. Taking closed circuits formed by each cell with the 8-ohm wire, we have, from the second law,

$$E_1 = 6 = 3I_1 + 8(I_1 + I_2) = 11I_1 + 8I_2$$

and

$$E_2 = 4 = 2I_2 + 8(I_1 + I_2) = 8I_1 + 10I_2.$$

Solving the two equations, we find $I_1 = \frac{14}{23} = 0.61$ amp, $I_2 = -\frac{2}{23} = -0.09$ amp.

The minus sign indicates that the current I_2 flows in the sense opposite to that shown in the diagram; i.e. it flows against the e.m.f. of the generator E_2 . It does so because the potential difference V_{AB} is greater than E_2 :

$$\begin{aligned} V_{AB} &= \left(R I_1 + I_2 \right) = 8 \left(\frac{14}{23} - \frac{2}{23} \right) \\ &= 8 \times \frac{12}{23} = \frac{96}{23} = 4.2 \text{ volts.} \end{aligned}$$

This is equal to the e.m.f. E_2 plus the drop across the internal resistance r_2 (p. 828):

$$\begin{aligned} V_{CD} &= 4 + 2 \times \frac{2}{23} = 4 + \frac{4}{23} \\ &= \frac{96}{23} \text{ volts} = V_{AB}. \end{aligned}$$

It is also equal to the e.m.f. E_1 minus the drop across r_1 , because the current flows through the upper generator in the direction of its e.m.f.:

$$\begin{aligned} V_{FH} &= 6 - 3 \times \frac{14}{23} = 6 - \frac{42}{23} = \frac{138 - 42}{23} \\ &= \frac{96}{23} \text{ volts} = V_{AB}. \end{aligned}$$

WHEATSTONE BRIDGE MEASUREMENT OF RESISTANCE

Wheatstone Bridge Circuit

About 1843 Wheatstone designed a circuit called a 'bridge circuit' which gave an accurate method for measuring resistance. We shall deal later with the practical aspects. In Fig. 33.28, X is the unknown resistance, and P , Q , R are resistance boxes. One of these—usually R —is adjusted until the galvanometer between A , C , represented by its resistance R_g , shows no deflection: that is to say,

$$I_g = 0.$$

Then, as we shall show,

$$\frac{P}{Q} = \frac{R}{X},$$

whence

$$X = \frac{Q}{P} R.$$

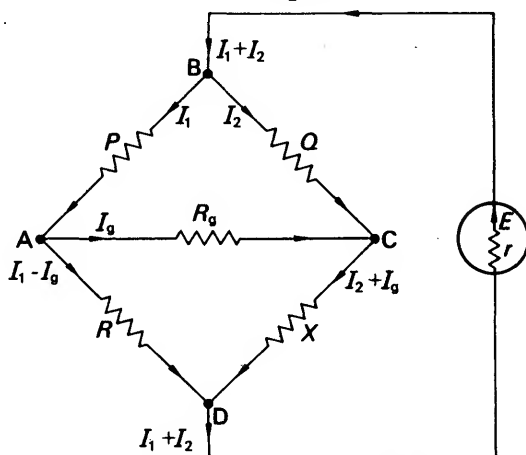


FIG. 33.28. Analysis of Wheatstone bridge.

Fig 33.28 shows Kirchhoff's first law applied to the circuit. From the second law, we have:

$$\text{loop ACBA: } R_g I_g - Q I_2 + P I_1 = 0, \quad (i)$$

$$\text{loop ACDA: } R_g I_g + X(I_2 + I_g) - R(I_1 - I_g) = 0,$$

$$\text{or } I_g(R_g + X + R) + X I_2 + R I_1 = 0. \quad (ii)$$

If we wished to find I_g , we would have to set up a third equation, by going round one of the loops, including the battery (p. 831). But if we wish only to find the condition for no deflection of the galvanometer, we have merely to put $I_g = 0$ in equations (i) and (ii). Then

$$-Q I_2 + P I_1 = 0, \quad \text{or} \quad P I_1 = Q I_2,$$

$$\text{whence} \quad \frac{P}{Q} = \frac{I_2}{I_1},$$

$$\text{and} \quad I X_2 - R I_1 = 0, \quad \text{or} \quad X I_2 = R I_1,$$

$$\text{whence} \quad \frac{R}{X} = \frac{I_2}{I_1}$$

Therefore, as already stated,

$$\frac{P}{Q} = \frac{R}{X}. \quad (19)$$

This is the condition for balance of the bridge. It is the same, as the reader may easily show, if the battery and galvanometer are interchanged in the circuit.

Alternative Wheatstone Bridge Proof

Equation (19) for the balance condition can be got without the use of Kirchhoff's laws. At balance, since no current flows through the galvanometer, the points A and C must be at the same potential (Fig. 33.29). Therefore

$$V_{AB} = V_{CB}$$

$$\text{and} \quad V_{AD} = V_{CD},$$

whence

$$\frac{V_{AB}}{V_{AD}} = \frac{V_{CB}}{V_{CD}}. \quad (i)$$

Also, since $I_g = 0$, P and R carry the same current, I_1 , and X and Q carry the same current, I_2 . Therefore

$$\frac{V_{AB}}{V_{AD}} = \frac{I_1 P}{I_1 R} = \frac{P}{R} \quad (ii)$$

and

$$\frac{V_{CB}}{V_{CD}} = \frac{I_2 Q}{I_2 X} = \frac{Q}{X}$$

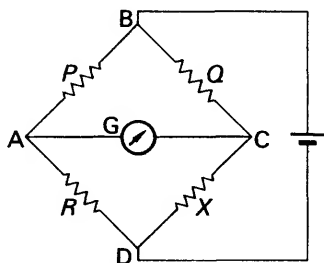


FIG. 33.29. Wheatstone bridge.

Hence by equations (i) and (ii),

$$\frac{P}{R} = \frac{Q}{X}$$

or

$$\frac{P}{Q} = \frac{R}{X}$$

Galvanometer Position

We shall now show, by taking a numerical example, how the galvanometer in a bridge circuit can best be positioned.

Fig. 33.30 shows an unbalanced Wheatstone bridge, fed from a cell

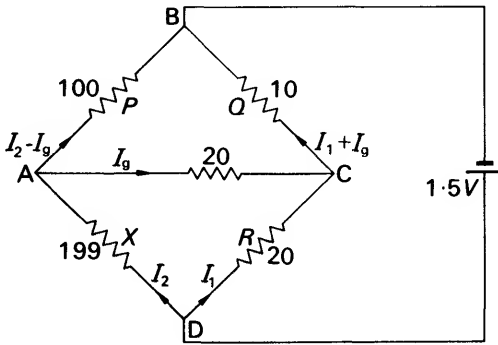


FIG. 33.30. An unbalanced Wheatstone bridge.

of negligible internal resistance. The figures give the resistance in ohms, and I_g is to be found. Applying Kirchhoff's laws:

$$\text{Loop ACBA: } 20I_g + 10(I_1 + I_g) - 100(I_2 - I_g) = 0$$

or

$$130I_g + 10I_1 - 100I_2 = 0,$$

whence

$$I_1 = 10I_2 - 13I_g. \quad (i)$$

$$\text{Loop ADCA: } -139I_2 + 20I_1 - 20I_g = 0.$$

$$\text{Substituting for } I_1: -199I_2 + 200I_2 - 260I_g - 20I_g = 0,$$

or

$$I_2 - 280I_g = 0,$$

whence

$$I_2 = 280I_g.$$

$$\therefore \text{by (i), } I_1 = 2800I_g - 13I_g = 2787I_g. \quad (ii)$$

$$\text{Loop DCBXD: } 20I_1 + 10(I_1 + I_g) = 1.5,$$

or

$$30I_1 + 10I_g = 1.5.$$

$$\text{Substituting from (ii) for } I_1: 30 \times 2787I_g + 10I_g = 1.5$$

or

$$83620I_g = 1.5,$$

whence

$$\begin{aligned} I_g &= \frac{1.5}{83620} = 1.79 \times 10^{-5} \text{ A} \\ &= 17.9 \text{ microamperes.} \end{aligned}$$

The reader should now show that, if the battery and galvanometer were interchanged, the current I_g would be 13.2 microamperes. This result illustrates an important point in the use of the Wheatstone bridge: with a given unbalance, the galvanometer current is greatest when the galvanometer is connected from the junction of the highest resistances to the junction of the lowest. Therefore, unless $P = Q$, which is unusual, the galvanometer should be connected across PQ .

Practical Arrangement

A practical form of Wheatstone bridge is shown in Fig. 33.31. The resistances P and Q can be given values of 10, 100, or 1000 ohms by three-point switches. The resistance R has four decade dials by which it can be varied from 1 ohm to more than 10000 ohms. Pairs of terminals are provided for connecting the unknown resistance, the battery, and the galvanometer, X, B, G ; and keys K_1 and K_2 are fitted in the battery and galvanometer circuits.

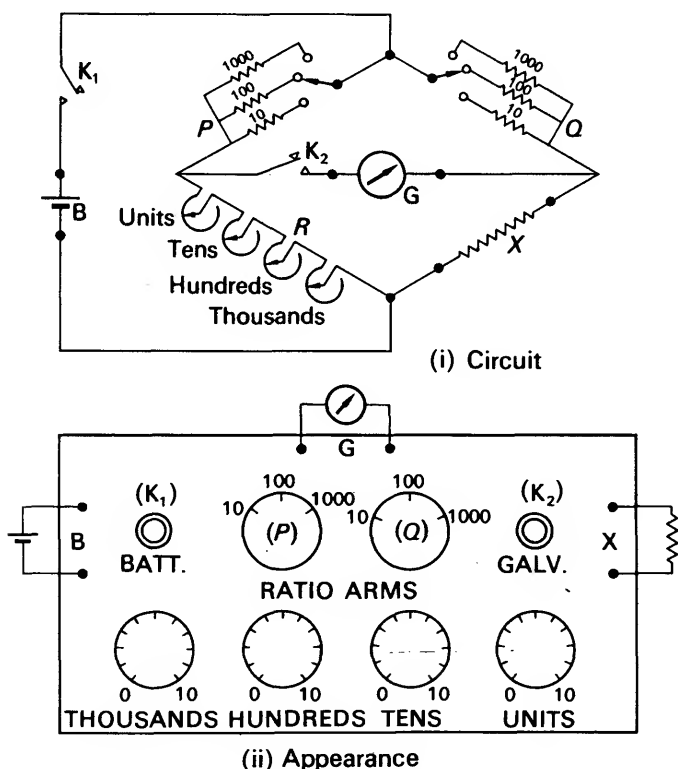


FIG. 33.31. Practical form of Wheatstone bridge.

To measure a resistance, we first set $P = Q = 10\ \Omega$. We set $R = 0$ and press first K_1 then K_2 ; the small interval between pressing K_1 and K_2 gives time for the currents in the bridge to become steady (Chapter

31). Let us suppose that, when we press K_2 , the galvanometer deflects to the right. We then set $R = 10000\ \Omega$ and again press K_1, K_2 . If the galvanometer deflects to the left we can proceed; if it deflects again to the right, then either we have made a wrong connexion—which with this form of bridge is almost impossible—or X is greater than $10000\ \text{ohms}$. If the galvanometer deflects to the left, we try again with $R = 1000\ \Omega$; and so on with $100\ \Omega$ and $10\ \Omega$, if necessary. Let us suppose that the galvanometer deflects to the left with $R = 100\ \Omega$, but to the right with $10\ \Omega$. We then adjust the 10 's dial until we get, say, a leftward deflection with $40\ \Omega$ and a rightward with $30\ \Omega$. With the unit's dial we now narrow the limits to, let us say, $36\ \Omega$ (left) and $35\ \Omega$ (right). We have

$$\frac{X}{R} = \frac{Q}{P} = \frac{10}{10} = 1.$$

$$\therefore X = R.$$

It follows that X lies between 35 and $36\ \Omega$.

We now set $P = 100\ \Omega$, so that

$$\frac{X}{R} = \frac{Q}{P} = \frac{10}{100} = \frac{1}{10}$$

or

$$X = \frac{R}{10}.$$

The balance-point now lies between $R = 350\ \Omega$ (right) and $R = 360\ \Omega$ (left); by using the unit's dial we can now locate it between, say, 353 and 354 . Then, from the equation, X lies between 35.3 and $35.4\ \text{ohms}$. If we finally make $P = 1000\ \Omega$, we have

$$\frac{X}{R} = \frac{Q}{P} = \frac{10}{1000} = \frac{1}{100}$$

or

$$X = \frac{R}{100}.$$

Only a sensitive galvanometer will give considerable deflections near the balance-point in this condition; if it locates the balance-point between $R = 3536\ \Omega$ and $3537\ \Omega$ then X lies between 35.36 and $35.37\ \text{ohms}$. If a moving-coil galvanometer is used, it must be protected by a high series resistance while balance is being sought.

Range of Measurable Resistance

The resistors P and Q are often called the ratio arms of the bridge, because their resistances determine the ratio of R to X . If X is greater than the greatest value of R , it can be measured by making $Q = 100$, $P = 10$. Then

$$\frac{X}{R} = \frac{Q}{P} = \frac{100}{10} = 10$$

and

$$X = 10R.$$

A balance-point between $R = 14620$ and 14630 , say, would mean that R lay between $146\ 200$ and $146\ 300$ ohms. Similarly, by making $Q = 1000$, $P = 10$, resistances can be measured up to 100 times the greatest value of R : that is to say, up to a little more than $1\ 000\ 000$ ohms. With these high resistances, however, the near-balance currents are very small, and a sensitive galvanometer is necessary.

The lowest resistance which a bridge of this type can measure with reasonable accuracy is about $1\ \text{ohm}$; R can be adjusted in steps of $1\ \text{ohm}$, and P/Q can be made $1/100$, so that measurements can be made to within $1/100\ \text{ohm}$. Resistances lower than about $1\ \text{ohm}$ cannot be measured accurately on a Wheatstone bridge, whatever the ratios available, or the smallest steps in R . They cannot because of the resistances of the wires connecting them to the X terminals, and of the contacts between those wires and the terminals to which they are, at each end, attached. This is the reason why the potentiometer method is more satisfactory for low resistances.

The Post Office Box

An old-fashioned type of Wheatstone bridge, with plugs instead of switches, is called the Post Office box, and is illustrated in Fig. 33.32. It is connected up and used in the same way as the dial type of bridge,

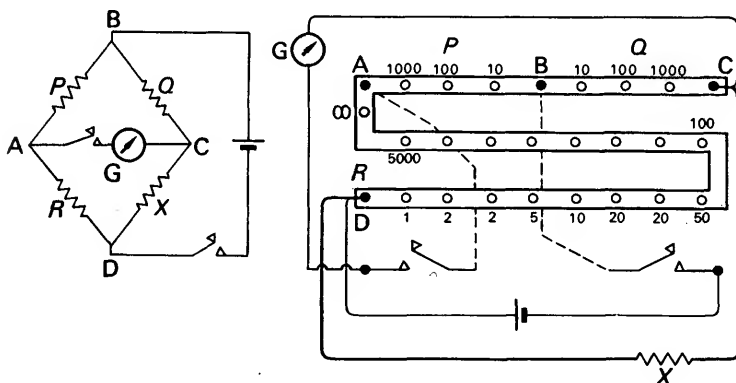


FIG. 33.32. Post Office box.

but requires far more skill by its operator. Anyone who has to use a Post Office box should observe the following rules:

- (i) do not attempt to memorize the wiring-up; the circuit should be worked out from the Wheatstone bridge diagram (Fig. 33.29);
- (ii) take the $10\ \Omega$ plugs out of each ratio arm P , Q before testing the circuit;
- (iii) test for correct connexions by seeing whether the galvanometer gives opposite deflections with $R = 0$ and $R = \infty$ (for the latter an 'infinity' plug is provided, whose gap is not bridged by any resistor—Fig. 33.32);

- (iv) press plugs home firmly, with a half-turn to the right ;
- (v) never mix the plugs from different boxes—always put loose plugs in the lid of their box, never on the bench.

The Slide-wire (Metre) Bridge

Fig. 33.33 shows a simple and cheap form of Wheatstone bridge; it is sometimes called a metre bridge, for no better reason than that the wire AB is often a metre long. The wire is uniform, as in a potentiometer, and can be explored by a slider S. The unknown resistance X and a known resistance R are connected as shown in the figure; heavy brass or copper strip is used for the connexions AD, FH, KB, whose resistances are generally negligible. When the slider is at a point C in

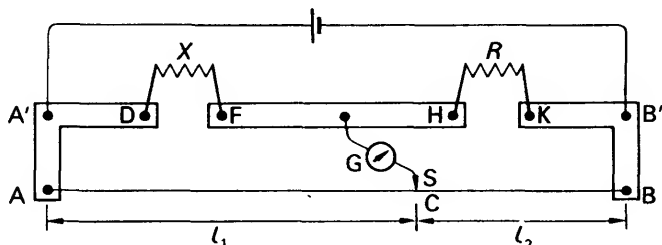


FIG. 33.33. Slide-wire (metre) bridge.

the wire it divides the wire into two parts, of resistances R_{AC} and R_{CB} ; these, with X and R , form a Wheatstone bridge. (The galvanometer and battery are interchanged relative to the circuits we have given earlier; that enables the slider S to be used as the galvanometer key. We have already seen that the interchange does not affect the condition for balance (p. 830).) The connexions are checked by placing S first on A , then on B . The balance-point is found by trial and error—not by scraping S along AB . At balance,

$$\frac{X}{R} = \frac{R_{AC}}{R_{CB}}.$$

Since the wire is uniform, the resistances R_{AC} and R_{CB} are proportional to the lengths of wire, l_1 and l_2 . Therefore

$$\frac{X}{R} = \frac{l_1}{l_2}. \quad (20)$$

The resistance R should be chosen so that the balance-point C comes fairly near to the centre of the wire—within, say, its middle third. If either l_1 or l_2 is small, the resistance of its end connexion AA' or BB' in Fig. 33.33 is not negligible in comparison with its own resistance; equation (20) then does not hold. Some idea of the accuracy of a particular measurement can be got by interchanging R and X , and balancing again. If the new ratio agrees with the old within about 1 per cent, then their average may be taken as the value of X .

Resistance by Substitution

Fig. 33.34 illustrates a simple way of measuring a resistance X . It is connected in series with a rheostat S , an ammeter A , and a cell. S is adjusted until A gives a large deflection. X is then replaced by a box of known resistances, R , which can be selected by plugs or dials. R is varied until the ammeter gives the same reading as before. Then, if the e.m.f. of the cell has not fallen, $R = X$. The accuracy of this method is limited, by the inherent error of the ammeter, to about 1 per cent. It does not depend on the accuracy of calibration of the ammeter, but on the accuracy with which it reproduces a given deflection for a given current. The ammeter is used simple as a 'transfer instrument'—to indicate when the current in the second part of the experiment is the same as in the first. This principle is very useful in measurements more difficult than that of resistance by direct method. For example, the power output of a small radio transmitter can be measured by making it light a lamp, which is placed near to a lightmeter (p. 571). The lamp is then connected to a source of direct current, and the current through it is adjusted until the light-meter gives the same reading. Simple measurements of the current, and the voltage across the lamp, then give the power supplied to it—which is equal to the power output of the radio transmitter.

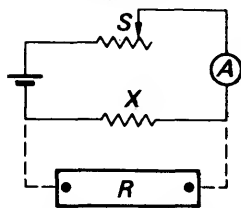


FIG. 33.34. Measurement of resistance by substitution.

Resistance and Sensitivity of a Galvanometer

It is often necessary to know the resistance, R_g , and sensitivity of a suspended-coil galvanometer. To find them, we may use the circuit of Fig. 33.35 (i). S is a rheostat of about 1000 ohms maximum, r is a standard shunt of about 0.01 ohm, M is a milliammeter, and R is a resistance box. The current in the main circuit, I , is adjusted to a value which can be accurately read on M —say 10 milliamperes. Since r is very

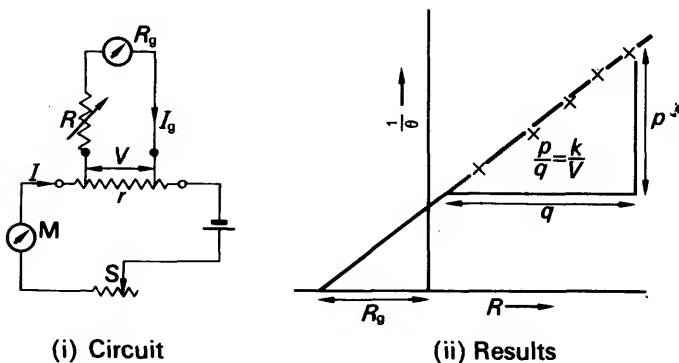


FIG. 33.35. Galvanometer calibration.

small compared with $R + R_g$ the galvanometer current I_g is negligible compared with I , and we may say that the potential difference across r is

$$V = rI.$$

In this example it would be $0.01 \times 10 \times 10^{-3} = 10^{-4}$ volt = 0.1 millivolt. The galvanometer current is therefore

$$I_g = \frac{V}{R + R_g}.$$

If $R + R_g$ is 1000 ohms, then $I_g = 10^{-4}/10^3 = 10^{-7}$ amp = 0.1 micro-ampere, which is a reasonable value for a galvanometer of moderate sensitivity. If θ is the deflection of the galvanometer, then

$$I_g = k\theta,$$

where k is a constant (the reduction factor) which we wish to find. From the equation for I_g above.

$$\frac{V}{R + R_g} = k\theta,$$

whence

$$\frac{1}{\theta} = \frac{k}{V}(R + R_g).$$

Therefore, if we vary R and plot $1/\theta$ against it, we get a straight line, as shown in Fig. 33.35 (ii). The line makes an intercept on the R axis, which gives the value of R_g . And its slope p/q is k/V . Since we know V from above, we can hence find k ; and $1/k$, the deflection per unit current, is the sensitivity.

Temperature Coefficient of Resistance

We have seen that the resistance of a given wire increases with its temperature. If we put a coil of fine copper wire into a water bath, and use a Wheatstone bridge to measure its resistance at various moderate temperatures t , we find that the resistance, R , increases

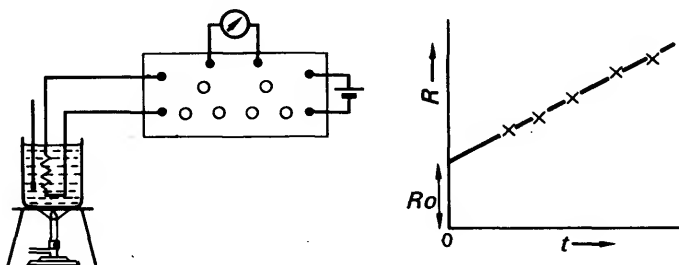


FIG. 33.36. Measurement of temperature coefficient.

uniformly with the temperature (Fig. 33.36). We may therefore define a temperature coefficient of resistance, α , such that

$$R = R_0(1 + \alpha t), \quad (21)$$

where R_0 is the resistance at 0°C . In words,

$$\alpha = \frac{\text{increase of resistance per deg. C rise of temperature}}{\text{resistance at } 0^\circ\text{C}}$$

If R_1 and R_2 are the resistances at $t_1^\circ\text{C}$ and $t_2^\circ\text{C}$, then

$$\frac{R_1}{R_2} = \frac{1 + \alpha t_1}{1 + \alpha t_2}. \quad (22)$$

Values of α for pure metals are of the order of 0.004 per deg. C. They are much less for alloys than for pure metals, a fact which enhances the value of alloys as materials for resistance boxes and shunts.

Equation (21) represents the change of resistance with temperature fairly well, but not as accurately as it can be measured. More accurate equations are given on p. 370 in the Heat section of this book, where resistance thermometers are discussed.

EXAMPLES

1. How would you compare the resistances of two wires A and B, using (a) a Wheatstone bridge method and (b) a potentiometer? For each case draw a circuit diagram and indicate the method of calculating the result.

In an experiment carried out at 0°C , A was 120 cm of Nichrome wire of resistivity 100×10^{-6} ohm cm and diameter 1.20 mm, and B a German silver wire 0.80 mm diameter and resistivity 28×10^{-6} ohm cm. The ratio of the resistances A/B was 1.20. What was the length of the wire B?

If the temperature coefficient of resistance of Nichrome is 0.00040 per deg. C and of German silver is 0.00030 per deg. C, what would the ratio of the resistances become if the temperature were raised by 100 deg. C? (L.)

First part (see pp. 830, 824).

Second part. With usual notation,

$$\text{for A,} \quad R_1 = \frac{\rho_1 l_1}{a_1},$$

$$\text{and for B,} \quad R_2 = \frac{\rho_2 l_2}{a_2}.$$

$$\therefore \frac{R_1}{R_2} = \frac{\rho_1}{\rho_2} \cdot \frac{l_1}{l_2} \cdot \frac{a_2}{a_1} = \frac{\rho_1}{\rho_2} \cdot \frac{l_1}{l_2} \cdot \frac{d_2^2}{d_1^2},$$

where d_2, d_1 are the respective diameters of B and A.

$$\therefore 1.20 = \frac{100}{28} \times \frac{120}{l_2} \times \frac{0.8^2}{10.2^2}.$$

$$\therefore l_2 = \frac{100 \times 120 \times 0.64}{1.20 \times 28 \times 1.44} = 159 \text{ cm} \quad (i)$$

When the temperature is raised by 100°C , the resistance increases according to the relation $R_t = R_0(1 + \alpha t)$. Thus

$$\text{new Nichrome resistance, } R_A = R_1(1 + \alpha \cdot 100) = R_1 \times 1.04,$$

$$\text{and new German silver resistance, } R_B = R_2(1 + \alpha' \cdot 100) = R_2 \times 1.03.$$

$$\therefore \frac{R_A}{R_B} = \frac{R_1}{R_2} \times \frac{1.04}{1.03} = 1.20 \times \frac{1.04}{1.03} = 1.21 \quad \quad \quad (\text{ii})$$

2. State Kirchhoff's laws relating to the currents in network of conductors. Two cells of e.m.f. 1.5 volts and 2 volts respectively and internal resistances of 1 ohm and 2 ohms respectively are connected in parallel to an external resistance of 5 ohms. Calculate the currents in each of the three branches of the network. (N.)

First part (see p. 828).

Second part. Suppose x , y amp are the respective currents through the cells (Fig. 33.37). Then, from Kirchhoff's first law, the current through the 5-ohm wire is $(x + y)$ amp.

Applying Kirchhoff's second law to the complete circuit with the cell of e.m.f. 1.5 volts and external resistance 5 ohms, we have

$$1.5 = x + 5(x + y) = 6x + 5y \quad \quad \quad (\text{i})$$

Applying the law to the complete circuit with the cell of e.m.f. 2 volts and external resistance 5 ohms, then

$$2 = 2y + 5(x + y) = 5x + 7y. \quad \quad \quad (\text{ii})$$

Solving (i) and (ii), we find $y = 9/34$ amp, $x = 1/34$ amp. Thus the current through the 5-ohm resistor $= x + y = 10/34 = 5/17$ amp.

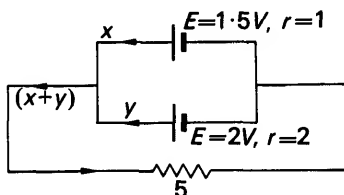


FIG. 33.37. Example.

EXERCISES 33

Circuit Calculations

1. State Ohm's law and describe an experiment to verify it.

A resistance of 1000 ohms and one of 2000 ohms are placed in series with a 100-volt mains supply. What will be the reading on a voltmeter of internal resistance 2000 ohms when placed across (a) the 1000 ohms resistance, (b) the 2000 ohms resistance? (L.)

2. Define the practical unit of potential difference and hence show that the rate of production of heat in a wire of constant resistance is proportional to the square of the current passing through it.

A cell A, e.m.f. 1.1 volts and internal resistance 3 ohms, is joined in parallel with another cell B, e.m.f. 1.4 volts and internal resistance 1 ohm, similar poles being connected together. The ends of a wire, of resistance 4 ohms, are joined to the terminals of A. Find (a) the current through the wire, (b) the rate of dissipation of energy in watts in each of the cells A and B. (L.)

3. Twelve cells each of e.m.f. 2 volts and of internal resistance $\frac{1}{2}$ ohm are arranged in a battery of n rows and an external resistance of $\frac{2}{3}$ ohm is connected to the poles of the battery. Determine the current flowing through the resistance in terms of n .

Obtain numerical values of the current for the possible values which n may take and draw a graph of current against n by drawing a smooth curve through the points. Give the value of the current corresponding to the maximum of the curve and find the internal resistance of the battery when the maximum current is produced. (L.)

4. Describe with full experimental details an experiment to test the validity of Ohm's law for a metallic conductor.

An accumulator of e.m.f. 2 volts and of negligible internal resistance is joined in series with a resistance of 500 ohms and an unknown resistance X ohms. The readings of a voltmeter successively across the 500-ohm resistance and X are $2/7$ and $8/7$ volts respectively. Comment on this and calculate the value of X and the resistance of the voltmeter. (N.)

5. State with reasons the essential requirement for the resistance of (a) an ammeter, (b) a voltmeter.

A voltmeter having a resistance of 1800 ohms is used to measure the potential difference across a 200 ohm resistance which is connected to the terminals of a d.c. power supply having an e.m.f. of 50 volts and an internal resistance of 20 ohms. Determine the percentage change in the potential difference across the 200 ohm resistor as a result of connecting the voltmeter across it. (N.)

6. State Ohm's law and describe how you would test its validity. Why would an experiment involving the use of a moving-coil ammeter and a moving-coil voltmeter be unsatisfactory?

In order to calibrate a galvanometer an accumulator of e.m.f. 200 volts and negligible resistance is connected in series with two resistances P and Q . A resistance R and the galvanometer, resistance G , are joined in series and then connected to the ends of P . The galvanometer is shunted by a resistance S . If $P = 200$ ohms, $Q = 1880$ ohms, $R = 291$ ohms, $S = 10$ ohms, $G = 90$ ohms, and the deflection is 20 divisions, calculate the current sensitivity in micro-amperes per division.

7. State Kirchhoff's laws for flow of electricity through a network containing sources of e.m.f.

Wheatstone bridge with slide wire of 0.5 ohm and length 50 cm is used to compare two resistances, each of 2 ohms. The cell has an e.m.f. of 2 volts and no internal resistance, and the galvanometer has resistance of 100 ohms. Find the current through the galvanometer when the bridge is 1 cm off balance. Compare the result with that of approximate calculation. (L.)

8. State Ohm's law, and describe the experiments you would make in order to verify it. The positive poles A and C of two cells are connected by a uniform wire of resistance 4 ohms and their negative poles B and D by a uniform wire of resistance 6 ohms. The middle point of BD is connected to earth. The e.m.f.s of the cells AB and CD are 2 volts and 1 volt respectively, their resistances 1 ohm and 2 ohms respectively. Find the potential at the middle point of AC. (O. & C.)

9. From the adoption of either the fundamental units cm, g, second, or the fundamental units m, kg, second, trace the steps necessary to define the volt and the ohm in terms of the ampere.

Discuss the suitability of (a) a moving-coil voltmeter, and (b) a slide-wire potentiometer for determining the potential differences in an experiment designed to verify Ohm's law.

Four resistors AB, BC, CD and DA of resistance 4 ohms, 8 ohms, 4 ohms and 8 ohms respectively are connected to form a closed loop, and a 6-volt battery of negligible resistance is connected between A and C. Calculate (i) the potential

difference between B and D, and (ii) the value of the additional resistance which must be connected between A and D so that no current flows through a galvanometer connected between B and D. (O. & C.)

Measurements

10. Describe and explain how you would use a potentiometer (a) to compare the electromotive forces of two cells, (b) to test the accuracy of the 1-amp reading of an ammeter.

The e.m.f. of a cell is balanced by the fall in potential along 150 cm of a potentiometer wire. When the cell is shunted by a resistance of 14 ohms the length required is 140 cm. What is the internal resistance of the cell? (N.)

11. Describe a metre bridge and the method of using it to compare the resistances of two conductors, giving the theory of the method. Why is the method unsatisfactory if the two resistances (a) differ widely from one another, (b) are very small?

A metre bridge is balanced with a piece of aluminium wire of resistance 7.30 ohm in the left-hand gap, the slide contact being 42.6 cm from the left-hand end of the bridge wire and the temperature 17°C . If the temperature of the aluminium wire is raised to 57°C , how may the balance be restored (a) by adjusting the slide contact, (b) by keeping the contact at 42.6 cm and connecting a conductor in parallel with the aluminium wire? (The temperature coefficient of resistance for aluminium may be taken as $3.8 \times 10^{-3} \text{ K}^{-1}$) (L.)

12. Describe and explain how a potentiometer is used to test the accuracy of the 1 volt reading of a voltmeter.

A potentiometer consists of a fixed resistance of 2030 ohms in series with a slide wire of resistance 4 ohm metre⁻¹. When a constant current flows in the potentiometer circuit a balance is obtained when (a) a Weston cell of e.m.f. 1.018 volt is connected across the fixed resistance and 150 cm of the slide wire and also when (b) a thermocouple is connected across 125 cm of the slide wire only. Find the current in the potentiometer circuit and the e.m.f. of the thermocouple.

Find the value of the additional resistance which must be present in the above potentiometer circuit in order that the constant current shall flow through it, given that the driver cell is a lead accumulator of e.m.f. 2 volt and of negligible resistance and the length of the slide wire is 2 metres. (L.)

13. Describe the Wheatstone bridge circuit and deduce the condition for 'balance'. State clearly the fundamental electrical principles on which you base your argument. Upon what factors do (a) the sensitivity of the bridge, (b) the accuracy of the measurement made with it, depend?

Using such a circuit, a coil of wire was found to have a resistance of 5 ohms in melting ice. When the coil was heated to 100°C , a 100 ohm resistor had to be connected in parallel with the coil in order to keep the bridge balanced at the same point. Calculate the temperature coefficient of resistance of the coil. (C.)

14. Explain the use of the Weston cell and the precautions to be taken when using it.

A steady current is passed through a manganin potentiometer wire AB of length 10 metres and diameter 0.56 mm connected to a resistance box BC with a resistance of 1001 ohms. A Weston cell of e.m.f. 1.018 volts, with a sensitive galvanometer in series with it, is connected in parallel between the points A and C. It is seen that no deflection is produced in the galvanometer. The Weston cell is removed and a thermocouple is connected *via* the galvanometer to points on the potentiometer wire 524 cm apart. Again there is no deflection. Draw the two

circuits and calculate the e.m.f. of the thermocouple. (Resistivity of manganin $= 41.87 \times 10^{-6}$ ohm cm.) (L.)

15. Describe, giving full experimental details, how you would compare the values of two unknown resistances each of the order of 0.1 ohm using a potentiometer. Draw a circuit diagram and give the theory of the method.

Suppose that, having set up the circuit to carry out this experiment, you found that no balance point could be obtained along the potentiometer. Discuss three possible reasons for this and the procedure you would adopt in order to trace it. (N.)

16. A two-metre potentiometer wire is used in an experiment to determine the internal resistance of a voltaic cell. The e.m.f. of the cell is balanced by the fall of potential along 90.6 cm of wire. When a standard resistance of 10 ohms is connected across the cell the balance length is found to be 75.5 cm. Draw a labelled circuit diagram and calculate, from first principles, the internal resistance of the cell.

How may the accuracy of this determination be improved? Assume that other electrical components are available if required. (N.)

17. Define *resistivity* and *temperature coefficient of resistance*.

Explain, with the help of a clear circuit diagram, how you would use a Post Office box to determine the resistance of an electric lamp filament at room temperature.

A carbon lamp filament was found to have a resistance 375 ohms at the laboratory temperature of 20°C. The lamp was then connected in series with an ammeter and a d.c. supply, and a voltmeter of resistance 1050 ohms was connected in parallel with the lamp. The ammeter and voltmeter indicated 0.76 amp and 100 volt respectively. The temperature of the carbon filament was estimated to be 1200°C. Estimate the mean value of the temperature coefficient of resistance of carbon between 20°C and 1200°C, and comment on your result. (L.)

18. Explain the principle of the potentiometer.

If a slide wire potentiometer of total resistance 5 ohms and a Weston standard cell of e.m.f. 1.0187 volt were available, together with the necessary auxiliary apparatus, describe in detail how you would (a) determine the variation of the e.m.f. of a thermocouple with the temperature difference between its junctions (maximum e.m.f. 3 millivolt), (b) compare the values of two resistances nominally 0.010 ohm and 0.005 ohm. (L.)

19. State Ohm's law. Deduce a formula for the resistance of a number of resistors connected in parallel.

If you were given a cell of constant internal resistance, an ammeter of negligible resistance and various wires of different lengths but otherwise identical, describe how you would determine how many cm of wire had the same resistance as the internal resistance of the cell.

Describe how you would compare the resistances of two resistors each about 10⁶ ohms using only an accumulator of e.m.f. of about 2 volts, a low-resistance uncalibrated galvanometer with a linear scale and full-scale deflection for about 10⁻⁷ amp and a three-terminal slider rheostat of about 20 ohms resistance. Justify the calculation of the resistance ratio from the readings you would take. (C.)

20. Describe how you would calibrate an ammeter using a standard resistor, a rheostat, accumulators, potentiometer slide wire with usual accessories, and a standard cell.

A standard low resistor is accidentally connected across a 100 volt d.c. main. It emits a momentary flash of light, vapourizes, and immediately breaks the

circuit without further sparking. Estimate the duration of the flash if the wire becomes incandescent at 550°C , melts at 900°C , has specific heat $0.42 \text{ kJ kg}^{-1} \text{ K}^{-1}$, density 5500 kg m^{-3} , is 10 cm long and has constant resistivity of 40 microhm cm. Assume that the rate of energy loss by the wire is small compared with the rate of heat development within it.

Give a physical explanation of the fact that this time is independent of the cross-sectional area of the wire. (C.)

chapter thirty-four

The Chemical Effect of the Current

In this chapter we shall deal both with the effects of an electric current when it is passed through a chemical solution, and with chemical generators of electric current, or cells.

ELECTROLYSIS

The chemical effect of the electric current was first studied quantitatively by Faraday, who introduced most of the technical terms which are now used in describing it. A conducting solution is called an *electrolyte* and the chemical changes which occur when a current passes through it are called *electrolysis* (*lysis* = decomposition). Solutions in water of acids, bases, and salts are electrolytes, and so are their solutions in some other solvents, such as alcohol. The plates or wires which dip into the electrolyte to connect it to the circuit are called *electrodes*; the one by which the current enters the solution is called the *anode*, and the one by which it leaves is called the *cathode* (Fig. 34.1 (i)).

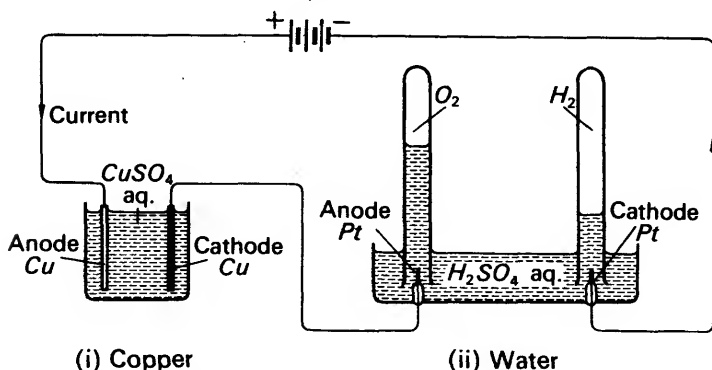


FIG. 34.1. Voltameters.

The whole arrangement is called a *voltameter*, presumably because it can be used to measure the current delivered by a voltaic cell; if the electrolyte is a solution of a copper or silver salt, the voltameter is called a copper or silver voltameter. If the electrolyte is acidulated water, then the voltameter is called a water voltameter, because when a current passes through it, the water, not the acid, is decomposed (Fig. 34.2 (ii)). We shall see why later.

Faraday's Laws of Electrolysis

When a current is passed through copper sulphate solution with copper electrodes, copper is deposited on the cathode and lost from the anode. Faraday showed that the mass dissolved off the anode by a given current in a given time is equal to the mass deposited on the cathode. He also showed that the mass is proportional to the product of the current, and the time for which it flows: that is to say, to the quantity of charge which passes through the voltameter. When he studied the electrolysis of water, he found that the masses of hydrogen and oxygen, though not equal, were each proportional to the quantity of charge that flowed. He therefore put forward his first law of electrolysis: *the mass of any substance liberated in electrolysis is proportional to the quantity of electric charge that liberated it.*

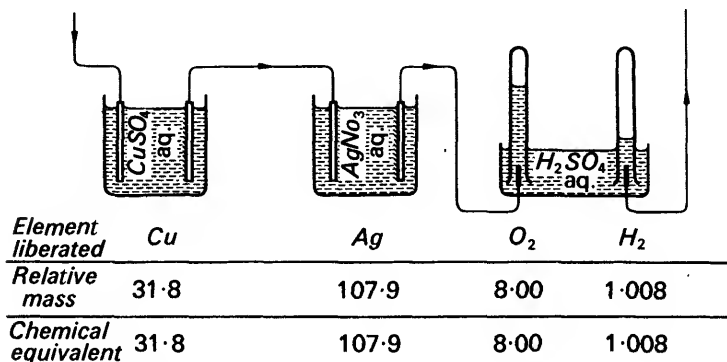


FIG. 34.2. Voltameters in series (same quantity of charge passes through each).

Faraday's second law of electrolysis concerns the masses of different substances liberated by the same quantity of charge. An experiment to illustrate it is in Fig. 34.2. The experiment shows that *the masses of different substances, liberated in electrolysis by the same quantity of electric charge, are proportional to the ratio of the relative atomic mass to the valency (also called the chemical equivalent).* This is Faraday's second law; it implies that the same quantity of charge is required to liberate one mole divided by the valency (also called the 'gramme-equivalent') of any substance. Recent measurements give this quantity as 96500 coulombs; it is called Faraday's constant, symbol F . It is also known as the *faraday*.

Electrochemical Equivalent

The mass of a substance which is liberated by one coulomb is called its *electrochemical equivalent*. It is expressed in kilogrammes per coulomb (kg C^{-1}) in SI units. If z is the electrochemical equivalent of a substance, the mass of it in kilogrammes liberated by I amperes in t seconds is

$$m = zIt. \quad (1)$$

Since the chemical equivalent of hydrogen is 1.008 then 1.008 g of hydrogen is liberated by 96500 coulombs, and the electrochemical equivalent of hydrogen is

$$z_{\text{H}} = \frac{1.008}{96,500} = 10.5 \times 10^{-5} \text{ g C}^{-1} = 1.05 \times 10^{-8} \text{ kg C}^{-1}$$

And similarly, since the chemical equivalent of copper is 31.8, its electrochemical equivalent is

$$z_{\text{Cu}} = \frac{31.8}{96,500} = 3.29 \times 10^{-4} \text{ g C}^{-1} = 3.29 \times 10^{-7} \text{ kg C}^{-1}$$

Measurement of Current by Electrolysis

In the past, the chemical effect of the current was used to define the ampere, because measurements with the current balance (p. 940) could not be made as accurately as simple weighings. Nowadays they can, and the ampere is defined as on p. 939. In those days the ampere was defined as the current which, when flowing steadily, would deposit 0.001118 g of silver per second.

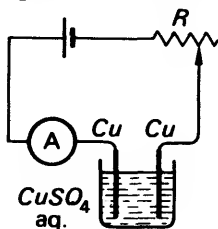


FIG. 34.3.
Current measurement
by chemical effect.

In the absence of a standard ammeter, the chemical effect can be used to find the error in a particular reading on an ammeter A (Fig. 34.3). A copper voltameter is connected in series with the ammeter, and a steady current passed for a known time t . The current is kept constant by adjusting the rheostat R to keep the deflection constant. The cathode is weighed before and after the experiment. Its increase in mass, m , gives the current I , in terms of the

electrochemical equivalent of copper, z :

$$m = zIt,$$

$$I = \frac{m}{zt} \quad (2)$$

The error in the ammeter is then the difference in the reading on A and the current calculated from (2).

Great care must be taken in this experiment over the cleanliness of the electrodes. They must be cleaned with sandpaper at the start; and, at the finish, the cathode must be rinsed with water and dried with alcohol, or over a gentle spirit flame: strong heating will oxidize the copper deposit.

The Mechanism of Conduction; Ions

The theory of electrolytic conduction is generally attributed to Arrhenius (1859–1927), although Faraday had stated some of its essentials in 1834. Faraday suggested that the current through an electrolyte was carried by charged particles, which he called ions (Greek

ion = go). A solution of silver nitrate, he supposed, contained silver ions and 'nitrate' ions. The silver ions were silver atoms with a positive charge; they were positive because silver was deposited at the cathode, or negative electrode (Fig. 34.4). The nitrate ions were groups of

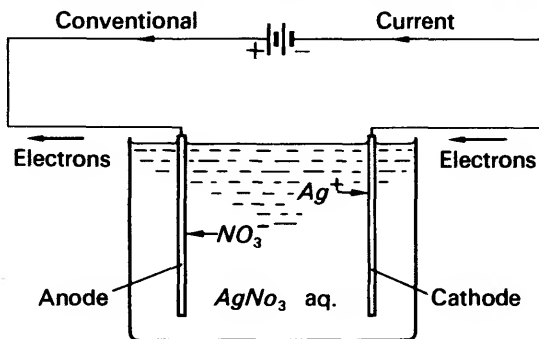


FIG. 34.4. Ions in electrolysis.

atoms— NO_3 groups—with a negative charge; they travelled towards the anode, or positive electrode, and, when silver electrodes are used, formed silver nitrate.

Nowadays, we consider that a silver ion is a silver atom which has lost an electron; this electron transfers itself to the NO_3 group when the silver nitrate molecule is formed, and gives the nitrate ion its negative charge. We denote nitrate and silver ions, respectively, by the symbols NO_3^- and Ag^+ . When the ions appear at the electrodes of a voltameter they are discharged. The current in the external circuit brings electrons to the cathode, and takes them away from the anode (Fig. 34.4). At the anode silver atoms lose electrons and go into solution as positive ions. *In effect*, the negative charges carried across the cell by the NO_3^- ions flow away through the external circuit. At the cathode, each silver ion gains an electron, and becomes a silver atom, which is deposited upon the electrode.

Ionization

The splitting up of a compound into ions in solution is called ionization, or ionic dissociation. Faraday does not seem to have paid much attention to how it took place, and the theory of it was given by Arrhenius in 1887. For a reason which we will consider later, Arrhenius suggested that an electrolyte ionized as soon as it was dissolved: that its ions were not produced by the current through it, but were present as such in the solution, before ever the current was passed.

We now consider that salts of strong bases and acids, such as silver nitrate, copper sulphate, sodium chloride, ionize completely as soon as they are dissolved in water. That is to say, a solution contains no molecules of these salts, but only their ions. Such salts are called strong electrolytes; so are the acids and bases from which they are formed, for these also ionize completely when dissolved in water.

Other salts, such as sodium carbonate, do not appear to ionize completely on solution in water. They are the salts of weak acids, and are called weak electrolytes. The weak acids themselves are also incompletely ionized in water.

Formation of Ions; Mechanism of Ionization

In the Heat section of this book we described the structure of the solid state (p. 294). In solid crystalline salts such as sodium chloride the structure is made up of sodium and chlorine ions: not of atoms, nor of NaCl molecules, but of Na^+ and Cl^- ions. In other words, we think today that ions exist in solid crystalline salts, as well as in their solutions. We do so for the reason that the idea enables us to build up a consistent theory of chemical combination, of the solid state, and of electrolytic dissociation.

A sodium atom contains eleven electrons, ten of which move in orbits close to the nucleus, and one of which ranges much more widely; for our present purposes we may represent it as in Fig. 34.5 (i). A chlorine atom has ten inner electrons and seven outer ones; for our present purposes we may lump these into two groups, as in Fig. 34.5 (ii). The outer electron of the sodium atom is weakly attracted to its nucleus, but the outer electrons of the chlorine atom are strongly attracted (because the ten inner electrons are a more effective shield round the $+11$ nucleus of sodium than round the $+17$ nucleus of chlorine). Therefore, when a sodium and a chlorine atom approach one another, the outer electron of the sodium atom is attracted more strongly by the chlorine nucleus than by the sodium nucleus. It leaves the sodium atom, and joins the outer electrons of

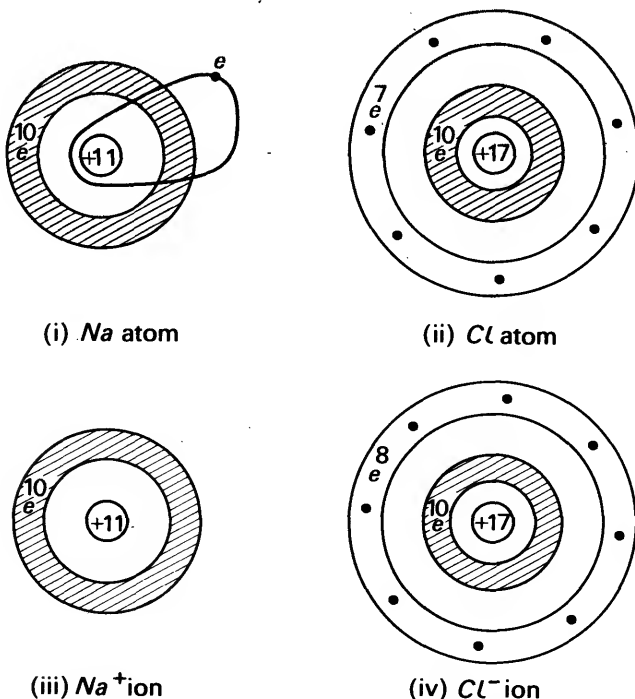
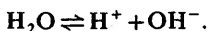


FIG. 34.5. Sodium and chlorine, atoms and ions.

the chlorine; the sodium atom becomes a positively charged sodium ion. Fig. 34.5 (iii), and the chlorine atom a negatively charged chlorine ion, Fig. 34.5 (iv). Between these two ions there now appears a strong electrostatic attraction, which holds them together as a molecule of NaCl. In the solid state, the ions are arranged alternately positive and negative; the forces between them bind the whole into a rigid crystal.

When such a crystal is dropped into water, it dissolves and ionizes. We can readily understand this when we remember that water has a very high dielectric constant: 81 (p. 774). It therefore reduces the forces between the ions 81 times, and the crystal falls apart into ions. In the same way we explain the ionization of other salts, and bases and acids. The idea that these dissociate because they are held together by electrostatic forces, which the solvent weakens, is supported by the fact that they ionize in some other solvents as well as water. These solvents also are liquids which have a high dielectric constant, such as methyl and ethyl alcohols (32 and 26 respectively). In these liquids, however, it seems that strong electrolytes behave as weak ones do in water: only a fraction of the dissolved molecules, not all of them, dissociate. In the electronic theory of atomic structure, the chemical behaviour of an element is determined by the number of its outer electrons. If it can readily lose one or two it is metallic, and forms positive ions; if it can readily gain one or two, it is acidic, and forms negative ions. Ions are not chemically active in the way that atoms are. Sodium atoms, in the form of a lump of the metal, react violently with water; but the hydroxide which they form ionizes ($\text{NaOH} \rightarrow \text{Na}^+ + \text{OH}^-$), and the sodium ions drift peaceably about in the solution—which is still mainly water.

Pure water is a feeble conductor of electricity, and we consider that it is but feebly ionized into H^+ and OH^- ions. These, we believe, are continually joining up to form water molecules, and then dissociating again in a dynamical equilibrium:



If, as we shall find in the electrolysis of water, H^+ and OH^- ions are removed from water, then more molecules dissociate, to restore the equilibrium.

The concentrations of H^+ and OH^- in water are so small that they do not contribute appreciably to the conduction of electricity when an electrolyte is dissolved in the water; but, as we shall see, they sometimes take part in reactions at the electrodes.

Explanation of Faraday's Laws

The theory of dissociation neatly explains Faraday's laws and some other phenomena of electrolysis. If an AgNO_3 molecule splits up into Ag^+ and NO_3^- ions, then each NO_3^- ion that reaches the anode dissolves one silver atom off it. At the same time, one silver atom is deposited on the cathode. Thus the gain in mass of the cathode is equal to the loss in mass of the anode. Also the total mass of silver nitrate in solution is unchanged; experiment shows that this is true. The mass of silver deposited is proportional to the number of ions reaching the cathode; if all the ions carry the same charge—a reasonable assumption—then the number deposited is proportional to the quantity of charge which deposits them. This is Faraday's first law.

To see how the ionic theory explains Faraday's second law, let us again consider a number of voltmeters in series (Fig. 34.6). When a

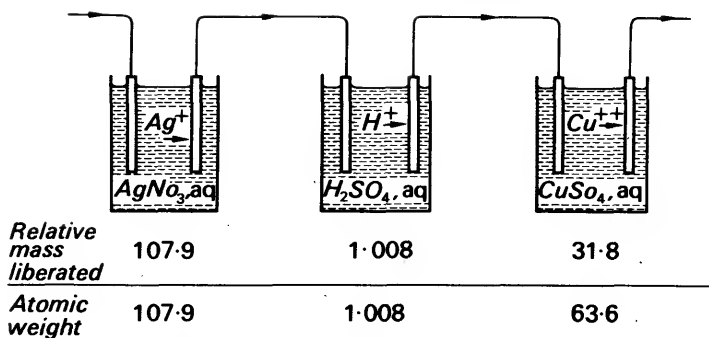


FIG. 34.6. Illustrating Faraday's second law.

current flows through them all, the same quantity of charge passes through each in a given time. Experiment shows that

$$\frac{\text{mass of silver deposited}}{\text{mass of hydrogen liberated}} = 107.0.$$

From experiments on chemical combination, we know that

$$\frac{\text{mass of silver atom}}{\text{mass of hydrogen atom}} = \frac{107.9}{1.008} = 107.0.$$

Therefore we may say that, each time a silver ion is discharged and deposited as an atom, a hydrogen ion is also discharged and becomes an atom. The hydrogen atoms thus formed join up in pairs, and escape as molecules of hydrogen gas. The theory fits the facts, on the simple assumption that the hydrogen and silver atoms carry equal charges: we now say that each is an atom which has lost one electron.

But when we consider the copper voltameter in Fig. 34.6, we find a complication. For

$$\frac{\text{mass of copper deposited}}{\text{mass of hydrogen liberated}} = \frac{31.8}{1.008},$$

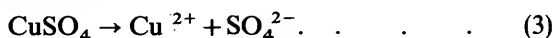
whereas

$$\frac{\text{mass of copper atom}}{\text{mass of hydrogen atom}} = \frac{63.6}{1.008}.$$

To explain this result we must suppose that only one copper atom is deposited for every two hydrogen atoms liberated. In terms of the ionic theory, therefore, only one copper ion is discharged for every two hydrogen ions. It follows that a copper ion must have twice as great a charge as a hydrogen ion: it must be an atom which has lost two electrons.

This conclusion fits in with our knowledge of the chemistry of copper. One atom of copper can replace two of hydrogen, as it does, for example, in the formation of copper sulphate, CuSO_4 , from sulphuric acid, H_2SO_4 . We therefore suppose that the sulphate ion also is doubly charged: SO_4^{2-} . When sulphuric acid is formed, two hydrogen atoms

each lose an electron, and the SO_4 group gains two. When copper sulphate is formed, each copper atom gives up two electrons to an SO_4 group. And when copper sulphate ionizes, each molecule splits into two doubly charged ions:



In general, if we express the charge on an ion in units of the electronic charge, we find that it is equal to the valency of the atom from which the ion was formed. That is to say, it is equal to the number of hydrogen atoms which the atom can combine with or replace.

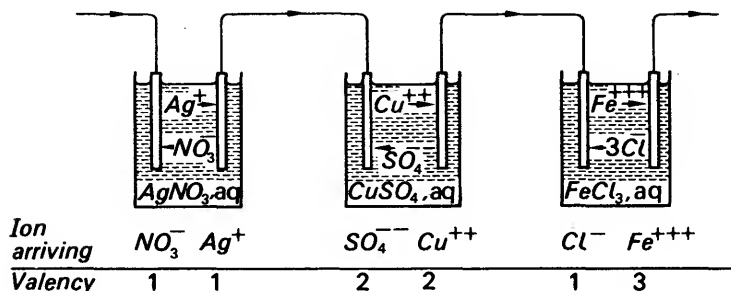


FIG. 34.7. Movement of ions in electrolysis.

This assertion, which is illustrated in Fig. 34.7, explains Faraday's second law (mass deposited \propto chemical equivalent). For if a current I passes through a voltameter for a time t , the total charge carried through it is It . And if q is the charge on an ion, the number of ions liberated is It/q . If M is the mass of an ion, the mass liberated is $M(It/q)$, and is therefore proportional to M/q . But M is virtually equal to the relative atomic mass, since the mass of an electron is negligible. And q , we have just seen, is equal, in electronic units, to the valency. Therefore the mass liberated is proportional to the ratio of relative atomic mass to valency. See also p. 845.

Electrolysis of Copper Sulphate Solution

If copper sulphate solution is electrolysed with platinum or carbon electrodes, copper is deposited on the cathode, but the anode is not dissolved away: instead, oxygen is evolved from it (Fig. 34.8). The SO_4^{2-} ions which approach the anode do not attack it; neither carbon nor platinum forms a sulphate, and each is said to be insoluble, in the electrolysis of copper sulphate. As the electrolysis proceeds, the solution becomes paler in colour; chemical tests show that it is gradually losing copper sulphate, but gaining sulphuric acid— Cu^{2+} ions are disappearing from the solution, but the SO_4^{2-} are remaining in it.

The oxygen which is evolved comes from the water of the solution. We have already seen that water is always slightly ionized, into H^+ and OH^- ions. When copper sulphate is electrolyzed with platinum or

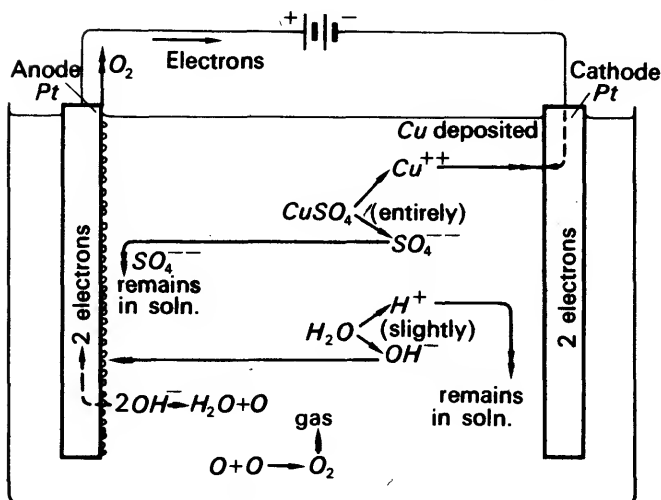
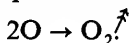


FIG. 34.8. Electrolysis of copper sulphate with insoluble electrodes.

carbon electrodes, the OH^- ions of the water are discharged at the anode. Each gives up an electron; they then combine in pairs to give a water molecule and an oxygen atom:



The oxygen atoms combine in pairs and come off as molecules:



As the OH^- ions disappear, an excess of H^+ ions appear in the solution. If the electrolysis were carried to the point where all the copper originally in the solution was deposited on the cathode, the solution would become simply one of sulphuric acid. This would be ionized into H^+ and SO_4^{2-} ions, in the proportion two H^+ to one SO_4^{2-} .

Electrolysis of Water

When a current is passed through water acidulated with dilute sulphuric acid, and platinum electrodes are used, oxygen and hydrogen are produced at the anode and cathode respectively. The amount of acid in solution remains unaltered, and the net effect is thus the electrolysis of water. The sulphuric acid ionizes into hydrogen and sulphate ions (Fig. 34.9):



The hydrogen ions from the acid greatly outnumber those from the water, but we cannot distinguish between them. All we can say is that, for every SO_4^{2-} ion that approaches the anode, two H^+ ions approach the cathode. At the cathode, the H^+ ions collect electrons, join up in pairs, and come off as molecules of hydrogen gas, H_2 . At the anode,

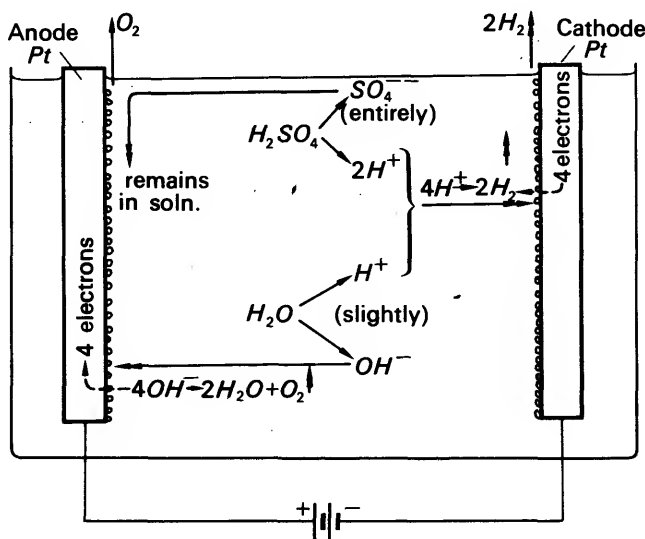


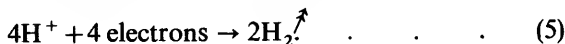
FIG. 34.9. Electrolysis of water.

however, the SO_4^{2-} ions remain in solution, and OH^- ions are discharged; as before, they form water and oxygen molecules, and the oxygen comes off as gas.

To produce one molecule of oxygen, four OH^- ions must be discharged:



And every time four OH^- ions are discharged at the anode, four H^+ ions are discharged at the cathode. (If this were not so, a net positive or negative charge would accumulate in the solution, and we could draw sparks from it.) Thus the reaction is accompanied by



Equations (4) and (5) agree with the experimental fact that hydrogen and oxygen come off in the proportions in which they are found in water: 2 to 1 by volume. This can be shown by collecting the gases in tubes filled with electrolyte and inverted over the electrodes (Fig. 34.1 (ii)). The electrolysis decomposes the water only, and leaves the acid unchanged. Equal numbers of H^+ and OH^- ions are discharged, and in the solution H^+ ions remain, in the proportion of two H^+ to one SO_4^{2-} .

The only function of the sulphuric acid is to increase the concentration of ions in the solution, and so to enable it to carry a greater current with a given potential difference than would pure water. The greater current discharges H^+ ions at a greater rate, and so causes the water to dissociate faster into H^+ and OH^- . Thus more ions are formed to carry the current.

The Electrolytic Capacitor

An electrolytic capacitor is one in which the dielectric is formed by electrolysis—by a secondary reaction at an insoluble electrode. It is made from two coaxial aluminium tubes, A and K in Fig. 34.10, with a solution or paste of ammonium borate between them. A current is passed through from A (anode) to K (cathode) and a secondary reaction at A liberates oxygen. The oxygen does not come off as a gas, however, but combines with the aluminium to form a layer of aluminium oxide over the electrode A. This layer is about $1/400$ cm thick, and is an insulator. When the layer has been formed, the whole system can be used as a capacitor, one of whose electrodes is the cylinder A, and the other the surface of the liquid or paste adjacent to A. Because the dielectric layer is so thin, the capacitance is much greater than that of a paper capacitor of the same size.

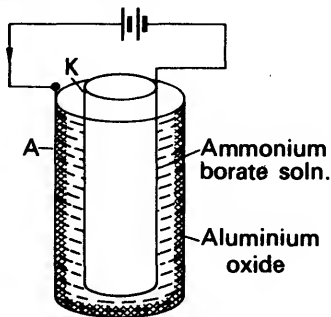


FIG. 34.10. Forming an electrolytic capacitor.

In the use of an electrolytic capacitor, some precautions must be taken. The voltage applied to it must not exceed a value determined by the thickness of the dielectric, and marked on the condenser; otherwise the layer of aluminium oxide will break down (see 'dielectric strength', p. 772). And the voltage must always be applied in the same sense as when the layer was being formed. If the plate A is made negative with respect to K, the oxide layer is rapidly dissolved away. Consequently an alternating voltage must never be applied to an electrolytic capacitor. This condition limits the usefulness of these capacitors.

Electrolytic capacitors are not very reliable, because the oxide layer is apt to break down with age. Domestic radio receivers abound in them, but in high-grade apparatus they are avoided.

Application of Ohm's Law to Electrolytes

Fig. 34.11 (i) shows how the current through an electrolyte, and the potential difference across it, may be measured. If the electrodes are soluble—copper in copper sulphate, for example—then the current is proportional to the potential difference (Fig. 34.11 (ii)). The best results in this experiment are obtained with very small currents. If the current is large, the solution becomes non-uniform: it becomes stronger near the anode, where copper is dissolved by the attack of the SO_4^{2-} ions, and weaker near the cathode, where copper is deposited, and SO_4^{2-} ions drift away. Near the cathode the solution becomes paler in colour; near the anode it becomes deeper. The total amount of copper sulphate in solution remains constant, but is gradually transferred to the neighbourhood of the anode. As the solution round the cathode becomes weaker, its resistance increases, and more than offsets the decreasing

resistance of the solution round the anode; with a given potential difference, therefore, the current gradually falls. A small current makes

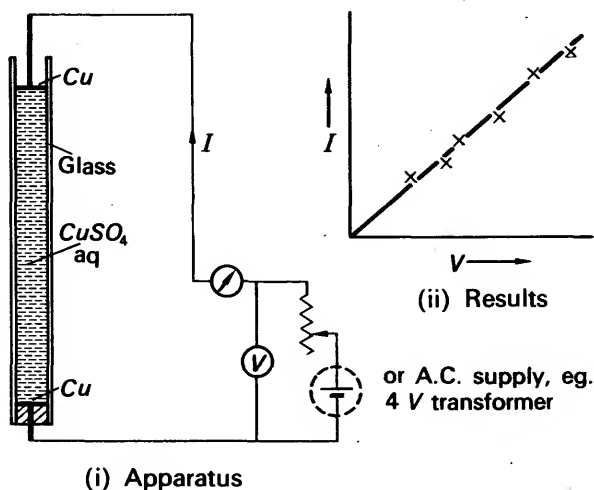


FIG. 34.11. Current/Voltage characteristic of electrolyte with soluble electrodes.

this effect unimportant, because it makes the electrolysis very slow. But the best way to do the experiment is to use an alternating current; the electrolysis reverses with the current fifty or more times per second, and no changes of concentration build up. Rectifier-type meters (p. 1011) are most suitable for measuring the current and potential difference in this case.

When the measurements are properly made they show, as we have said, that the current is proportional to the potential difference: for example, copper sulphate solution, with copper electrodes, obeys Ohm's law. The voltmeter behaves as a passive resistor; all the electrical energy delivered to it by the current appears as heat (pp. 791–2); no electrical energy is converted into mechanical or chemical work. In particular, therefore, no electrical energy is used to break up the molecules of copper sulphate into ions. This is the argument which led Arrhenius to suggest that the electrolyte dissociates into ions as soon as it is dissolved; dissociation is a result of solution, not of electrolysis.

Measurement of Resistance of Electrolyte

The resistance of an electrolyte can be measured on a Wheatstone bridge, preferably with an alternating current supply (Fig. 34.12). A telephone earpiece T is used as the detector, in place of the galvanometer—it gives minimum sound at the balance-point. 50 Hz mains

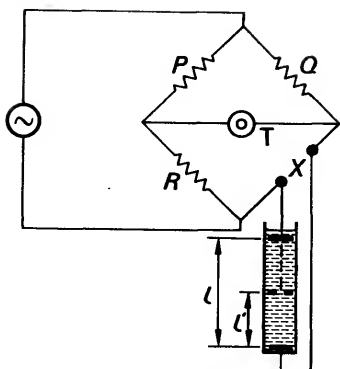


FIG. 34.12. Measurement of electrolyte resistance.

give an uncomfortably low-pitched note for listening, and a supply of frequency 400 to 1000 Hz is more satisfactory. This may be a small induction coil of a type sold for the purpose, which has fewer secondary turns than the common spark-coil, because a high voltage is not required. By sliding the upper electrode of the cell, the resistances of two lengths of electrolyte, l and l' are measured; their difference, r , is the resistance of a length $l-l'$, and is free from end-errors. If A is the cross-section of the tube, then the specific resistance of the electrolyte, ρ , is

given by

$$r = \frac{(l-l')\rho}{A}.$$

It is usual, however, not to give the resistivity of an electrolyte, but to give instead its conductivity, σ . This is defined as the reciprocal of its resistivity, and is expressed in mho per centimetre:

$$\sigma = \frac{1}{\rho} = \frac{l-l'}{Ar}.$$

Electrical Behaviour with Insoluble Electrodes

If we set out to find the current/voltage relationship for a water voltameter, using a d.c. supply, we find that the voltameter does not obey Ohm's law. If we apply to it a voltage E less than 1.7 volts, the current flows only for a short time and then stops, as though the voltameter were charged like a capacitor. To study the matter further, we may arrange a two-way key and a second galvanometer, G , as in Fig. 34.13. We first press the key at Y , and pass a current through the voltameter until it stops; then we press the key at X , and connect the galvanometer G straight across the voltameter. A brief current I flows through G , whose direction shows that the anode of the voltameter is acting as the positive pole of a current supply. It appears, therefore, that the voltameter is setting up a back-e.m.f. which prevents a steady current from flowing, unless the supply voltage is greater than 1.7 volts.

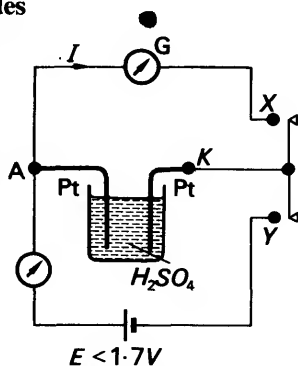
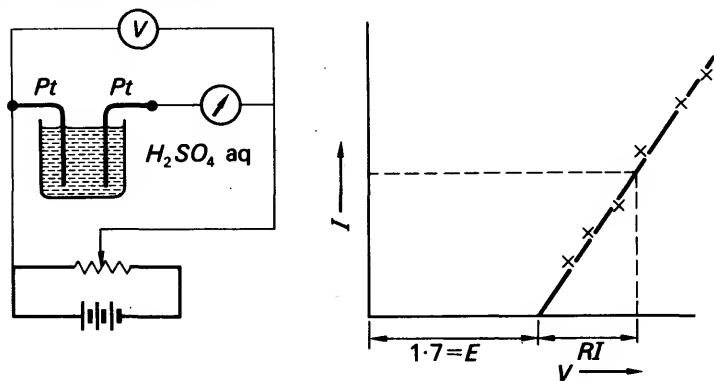


FIG. 34.13.
Demonstrating back-e.m.f. of water voltameter.

To investigate this behaviour further, let us gradually increase the potential difference across the voltmeter, as in Fig. 34.14 (i). We then



(i) Apparatus

(ii) Results

FIG. 34.14. Current/Voltage characteristic of water voltmeter.

find that, if we work with small currents, the current increases linearly with the potential difference V , when the latter is greater than 1.7 volts (Fig. 34.14 (ii)). It follows that the voltmeter does indeed exert a back-e.m.f., E , equal to 1.7 volts. If V is the potential difference applied to it, then the current through it obeys the relationship

$$I \propto (V - E).$$

We may write this as

$$I = \frac{V - E}{R},$$

where R is the resistance of the electrolyte.

Electrical Energy Consumed in Decomposition

We have not yet explained the origin of the back-e.m.f. E , but we shall try to do so later. Meanwhile let us notice that the behaviour of the voltmeter is somewhat like that of an electric motor. When the armature of a motor rotates, a back-e.m.f. is induced in it, and the current through it is given by an equation similar to that above. The back-e.m.f. in the motor, we shall see, represents the electrical power converted into mechanical work. So here the back-e.m.f. E represents this electrical power converted into chemical work—used in breaking up the water molecules. The potential difference across the voltmeter is

$$V = IR + E$$

from the equation for I ; the power equation is therefore

$$IV = I^2R + EI.$$

The left-hand term is the electrical power input; the first term on the right is the heat produced per second in the electrolyte; and the second term is the work done per second in decomposing the water.

Chemists tell us that when oxygen and hydrogen combine to form one mole of water (18 g) then 286 000 joules of heat are evolved. The energy set free is therefore 286 000 joules.

When one mole of water is decomposed, this much work must be done. In the process two moles of hydrogen are liberated (because the formula for water is H_2O). The quantity of electricity required to decompose one mole of water is therefore 2×96500 coulombs. If the back-e.m.f. is E , the corresponding amount of energy is $2 \times 96500 \times E$ joules. Therefore

$$2 \times 96500 \times E = 286\,000$$

whence

$$E = \frac{286\,000}{2 \times 96500} \\ = 1.48 \text{ volts.}$$

The lowest value of E which anyone has ever got by experiment is 1.67 volts—from which it appears that we have something yet to learn about what happens in a water voltameter.

CELLS

If we put plates of copper and zinc into a beaker of dilute sulphuric acid, we have a voltaic cell (p. 785). It is often called a simple cell. If we join its plates via a galvanometer, current flows through the galvanometer from the copper to the zinc (Fig. 34.15); the cell sets up an e.m.f. which acts, in the external circuit, from copper to zinc. Its value is about one volt. The copper plate is at a higher potential than the zinc plate, and is the positive terminal of the cell; the zinc is the negative terminal. Within the cell there must be some agency which

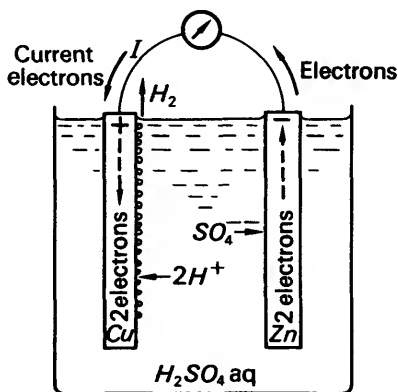


FIG. 34.15. A simple cell.

carries the current from the zinc to the copper. This is the agency which gives rise to the e.m.f. of the cell; it is analogous to the force exerted by the magnetic field on the moving electrons in the armature winding of a dynamo.

In a voltaic cell, the agency which gives rise to the e.m.f. is not so easy to track down as in a dynamo; there has been much argument about what is called the 'seat' of the e.m.f. We may start to seek it by placing a penknife blade into a strong solution of copper sulphate: a pink film of metallic copper is deposited on the blade. It appears, then, that copper ions have a tendency to go out of solution on to iron. In the same way we can show that they tend to go out on to zinc.

Do metal ions ever tend to go the other way—from solid metal into solution? They certainly do if the solution is sulphuric acid and the metal zinc or iron: the metal enters the solution in the form of ions, displaces the hydrogen ions, which are discharged and come off as gas, and turns the solution into one of iron or zinc sulphate. But this happens only if the zinc or iron is impure. Pure zinc in sulphuric acid gives no action at all—no zinc sulphate, no hydrogen.

Action in a Simple Cell

If we want to make pure zinc react with sulphuric acid, we must make it into part of a voltaic cell: we must connect it to a plate of a different metal, such as copper which also dips in the acid. Then the zinc is eaten away, and hydrogen bubbles off; but the hydrogen appears at the copper plate, and not at the zinc (Fig. 34.15). At the same time the solution becomes one of zinc sulphate—which simply means that it contains zinc ions in place of hydrogen ones. We can now form a picture of what happens when a stick of pure zinc is put alone into dilute sulphuric acid (Fig. 34.16). At first zinc ions leave the metal, and go into the liquid. But they leave negative charges on the zinc rod, which attract the zinc ions, and prevent any more from leaving. Nothing further happens. But if we introduce a plate of copper, and connect it to the zinc, electrons can flow from the zinc to the copper (Fig. 34.15). At the copper they can neutralize the charges on hydrogen ions, and to enable molecules to form and come off in bubbles. As the electrons flow away from the zinc, more zinc ions can go into solution, and so the zinc can continuously dissolve in the acid. In doing so, it maintains a continuous electric current in the wire connecting it to the copper. (When the zinc is impure each speck of impurity acts as the other plate of a minute cell, and enables the zinc around it to react with the acid. This 'local action', as it is called, makes impure zinc undesirable in voltaic cells, for it consumes the zinc without giving any useful current. It can be prevented by rubbing the zinc with mercury, which dissolves it and presents a surface of pure zinc to the acid. The process is called amalgamating the zinc.)

To explain the voltaic cell, therefore, we must suppose that zinc ions tend to dissolve

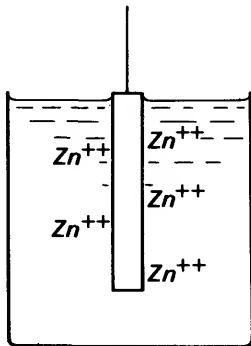


FIG. 34.16.
Pure zinc in dilute
sulphuric acid.

from zinc into sulphuric acid, but copper ions do not. This is consistent with the fact that copper does not react chemically with cold dilute sulphuric acid. The passage of ions from metals to solutions, and oppositely, was studied by Nernst about 1889; we shall give a slight account of his theory later.

Daniell's Cell

Fig 34.17 shows a cell, developed by Daniell about 1850, which has some advantages over Volta's. Daniell's cell consists of a zinc rod, Zn, in a porous pot, P, containing sulphuric acid; this in turn stands in a strong solution of copper sulphate in a copper vessel, Cu. (Sometimes the copper is just a thin sheet in a glass vessel.) When the copper and

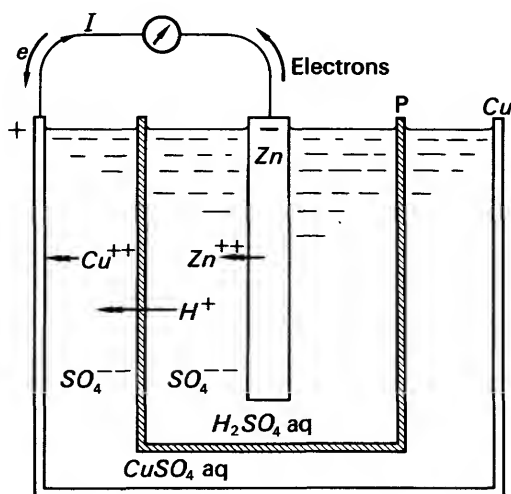


FIG. 34.17. A Daniell cell.

zinc are connected by a wire, current flows through the wire from copper to zinc. The copper is therefore the positive terminal of the cell, and the zinc the negative. As in Volta's cell, zinc ions go into solution at the zinc rod, leaving electrons on it. But at the copper plate, copper ions go out of solution—a metallic film of copper is deposited on the vessel. When the zinc and copper are joined by a wire, electrons from the zinc can go along it to the copper vessel, and discharge the copper ions as they reach it. To complete the action of the cell hydrogen ions from the sulphuric acid pass through the porous pot into the copper sulphate solution (Fig 34.17). Thus zinc is dissolved, the acid gradually changes to zinc sulphate, the copper sulphate gradually changes to sulphuric acid, and copper is deposited on the copper vessel.

The e.m.f. of a Daniell cell is about 1.08 volts. Its internal resistance depends on its size and condition—the size is usually about that of a plant-pot, and the internal resistance is of the order of several ohms.

Polarization

The great disadvantage of the simple cell is that it does not give a steady current; from the moment of making the circuit, the current starts to fall, and after a minute or two it almost ceases to flow. The current decays because a layer of hydrogen gas forms over the copper plate; scraping the plate enables the current to start once more, but it soon decays again. The hydrogen layer increases the internal resistance of the cell, but we do not believe that this is the main reason for the decay of the current. If the copper plate is replaced by one of platinum black (platinum with a finely grained surface), bubbles form on it very easily, and escape readily. The hydrogen layer may then be no more than one molecule thick; but the current decays as before.

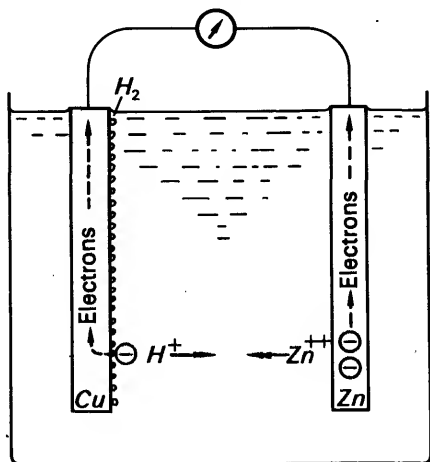


FIG. 34.18. Polarization in simple cell.

To explain this decay of the current we think that the hydrogen layer replaces the copper as an electrode of the cell. We suppose that the hydrogen tends to go back into solution as positive hydrogen ions (Fig. 34.18). In other words, it tends to behave in the same way as the zinc rod, on the other side of the cell, which also goes into solution as positive ions. Thus the hydrogen sets up an e.m.f. which opposes the original e.m.f. of the cell, and cuts down the current; the hydrogen thus sets up a 'back-e.m.f.' in the circuit. This behaviour is called polarization of the cell.

Depolarization

The advantage of Daniell's cell over Volta's is that it does not polarize. Hydrogen ions drift from the acid compartment into the copper sulphate compartment, but they are never discharged, no hydrogen molecules are formed, and no layer of hydrogen appears on the copper electrode. The copper sulphate solution is often called the depolarizer, because it prevents the formation of hydrogen gas.

Polarization in Water Voltmeter

We can find support for the idea of polarization in the behaviour of the water voltmeter (p. 856). When the potential difference across the voltmeter is less than 1.7 volts, the current through it falls to zero in a minute or less. The voltmeter itself can then deliver a current for a short time. Its positive terminal, as a source, is that which was its anode, and its negative terminal is that which was its cathode. While current was being sent through the voltmeter, the cathode became covered with hydrogen, and the anode with oxygen. When the voltmeter acts as a source of current it has, in effect, electrodes of oxygen and hydrogen and the current through the external circuit flows from the oxygen plate to the hydrogen (Fig. 34.19). In the simple cell, when it is polarized, there is no oxygen plate, but there is a hydrogen one, and this lies over the copper plate. We may therefore suppose

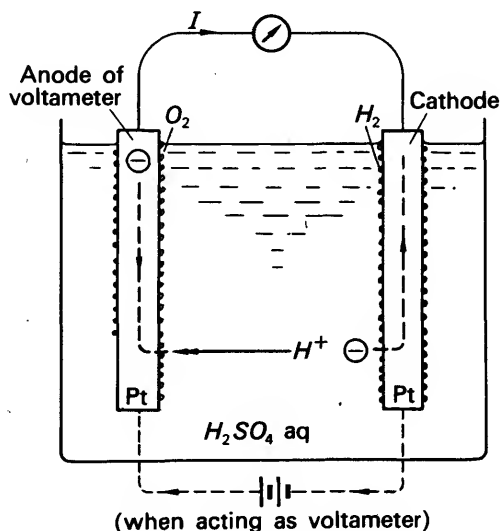


FIG. 34.19. Polarization of water voltmeter.

that it tends to drive a current, through the external circuit, from zinc to copper; that is to say, it sets up an e.m.f. opposing that of the cell with the copper plate clean.

Nernst's Theory of the Voltaic cell; Electrode Potentials

If a metal is in contact with a solution of one of its own salts, it is surrounded by its own ions. Whether the ions deposit themselves on the metal, or the metal goes into solution, depends partly on the particular metal concerned, and partly on the strength of the solution: the stronger the solution the greater its tendency to deposit ions on the metal. If the solution deposits ions, the metal comes to a positive potential with respect to it; if the metal goes into solution as ions, it

becomes negative with respect to the solution (Fig. 34.20 (i)). By methods beyond the scope of this book, the potential difference between a metal and a solution can be measured. These show that copper in normal copper sulphate solution (1 g-equivalent weight per litre) becomes 0.08 volt positive with respect to the solution.

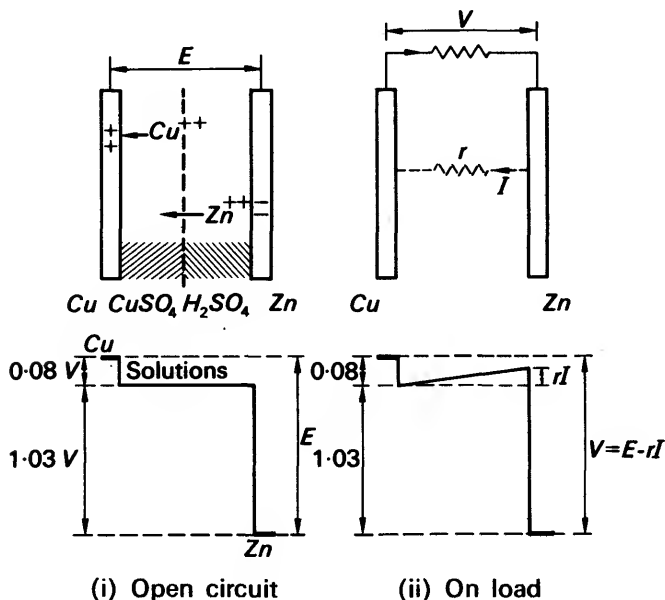


FIG. 34.20. Potential differences in voltaic cell.

Zinc in normal zinc sulphate solution becomes 1.03 volts negative. Now let us imagine a Daniell cell in which zinc sulphate replaces the sulphuric acid, and both solutions are normal. If we suppose that the solutions themselves set up no appreciable potential difference at their interface, then we get a potential distribution like that shown in Fig. 33.20 (i). The difference in potential between the copper and zinc is very nearly equal to the e.m.f. of a Daniell cell: 1.11 volts compared with 1.08. We may attribute the difference to the fact that a Daniell cell in practice has sulphuric acid, not zinc sulphate solution, in contact with the zinc; also the solutions are not normal: the acid is usually 1 to 4 of water, and the copper sulphate is saturated.

These considerations explain a striking experimental fact about all cells: the e.m.f. depends only on the nature and concentration of the constituent chemicals. The size of a cell affects only its internal resistance.

When a current I is drawn from a cell, there is a voltage drop across the internal resistance r , that is to say, across the solution or solutions. This modifies the potential diagram as shown in Fig. 34.20 (ii). The terminal voltage V , which is the observed potential difference between the copper and zinc, is now less than its open-circuit value, which is the e.m.f. of the cell.

The Leclanché Cell

Daniell's cell has the great practical disadvantage that it cannot be left set up; the solutions gradually mix by diffusion through the porous pot. It is now used only in teaching laboratories as a simple standard of e.m.f.; its e.m.f. is more nearly constant than that of any other

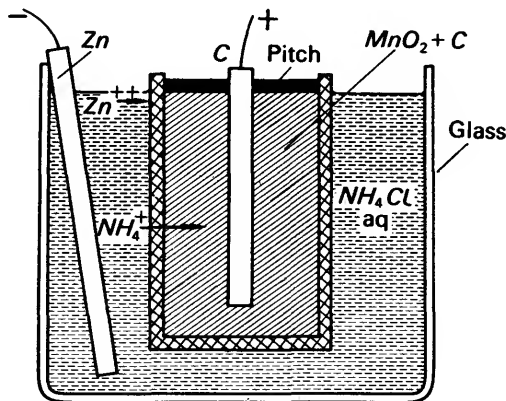


FIG. 34.21. Leclanché cell.

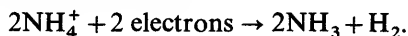
cheap and robust type of cell, and it is remarkably free from polarization.

A practically more useful cell is that devised by Leclanché. Its negative electrode is a zinc rod in a strong solution of ammonium chloride (Fig. 34.21). Its positive electrode is a carbon plate in a porous pot packed with manganese dioxide, which acts as the depolarizer. Manganese dioxide is a poor conductor of electricity, and powdered carbon is therefore packed in the pot with it.

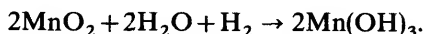
The ammonium chloride ionizes into ammonium ions and chlorine ions:



The zinc goes into solution, as zinc ions, and the ammonium ions drift through the porous pot towards the carbon plate. When a current is drawn from the cell, electrons flow from the zinc to the carbon, and discharge the NH_4^+ ions. The chemical action in this cell is complicated, but may be crudely represented as



The hydrogen tends to polarize the cell, but is gradually oxidized by the manganese dioxide; again the action is complicated, but it reduces to



The depolarizing action is slow, and a Leclanché cell is therefore not suitable for giving a large current for a long time. A short rest, however, enables the manganese dioxide to remove the hydrogen and restore

the e.m.f. of the cell. Thus Leclanché cells are suitable for giving intermittent currents: they are widely used, for example, with electric bells. They are also suitable for Wheatstone bridges, because they cannot give a current large enough to burn out the resistance coils when a wrong connexion is made. The e.m.f. of a Leclanché cell, before polarization sets in, is about 1.5 volts, and its internal resistance about 1 ohm.

Dry Cell

Fig. 34.22 shows a dry form of Leclanché cell, which has the obvious advantage that it is portable. The ammonium chloride is made into a paste with water, zinc chloride, flour, and gum; and the porous pot is replaced by a muslin bag. A cardboard spacer prevents the bag from touching the zinc and short-circuiting the cell. A dry Leclanché cell has the same e.m.f. as a wet one, but, for a given size, a lower internal resistance, because the thickness of solution between the zinc and carbon is less. It depolarizes better, because the volume of manganese dioxide is greater in relation to the overall size of the cell: the cycle-lamp size will give a useful light continuously for two hours or more.

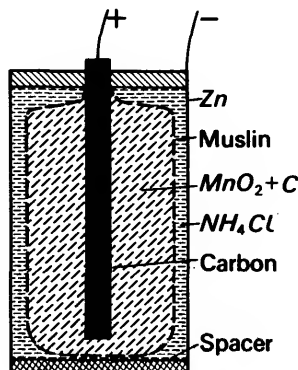


FIG. 34.22. A dry cell.

Standard Cells

A standard cell is one whose e.m.f. varies very little with time, and with temperature, so that it can be used as a standard of potential difference in potentiometer experiments. The commonest type is the Weston cadmium cell (Fig. 34.23). It is housed in an H-shaped glass

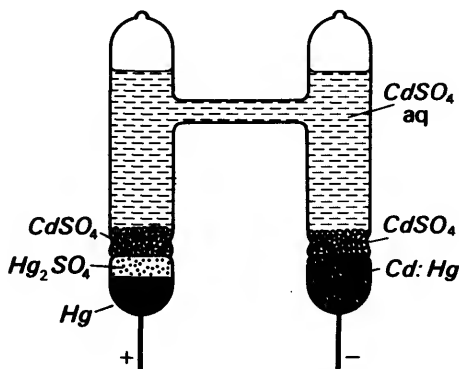


FIG. 34.23. A Weston cadmium cell.

tube because its electrodes are liquid or semi-liquid. The negative electrode is an amalgam of cadmium in mercury; the solution is of cadmium sulphate; the depolarizer is a paste of mercurous sulphate; and the positive electrode is mercury. In some cells, crystals of cadmium sulphate are placed on top of the electrodes to keep the solution saturated. The e.m.f. of one of these, in volts at a temperature $t^{\circ}\text{C}$, is

$$E = 1.01830 - 0.0000406(t - 20) + 0.00000095(t - 20)^2 + 0.00000001(t - 20)^3.$$

The e.m.f. of the type without crystals is about 1.0186 volts between 0°C and 40°C .

A standard cell without crystals of cadmium sulphate is called an unsaturated cell; one with crystals is called a saturated cell, because the crystals keep the solution saturated. Saturated cells give an accurately reproducible e.m.f., because the concentration of the solution is sharply defined at any given temperature. Unsaturated cells do not agree among one another so well, because the solution may vary a little from one to the other. But the e.m.f. of a given unsaturated cell varies less with temperature than that of a saturated cell, because the concentration of the solution is constant.

The depolarizer of a standard cell is effective only for very small currents, and the e.m.f. of the cell will change appreciably if more than about 10 microamperes are drawn from it. *A standard cell must not, in any circumstances, be used as a source of current.* In the early stages of balancing a standard cell against a potentiometer wire, a protective resistance of about 100 000 ohms should be connected in series with the cell.

Primary and Secondary Cells

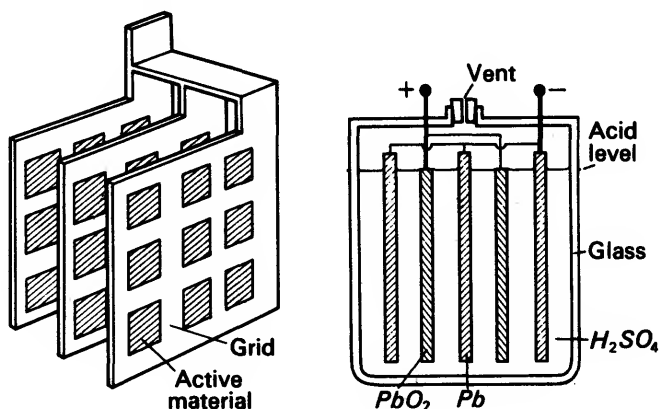
The cells which we have so far described are called primary cells. When they are run-down, their active materials must be renewed; the cells cannot be recharged by passing a current through them from another source. A secondary cell is one which can be recharged in this way.

SECONDARY CELLS

The Lead Accumulator

The commonest secondary cell is the lead-acid accumulator. Its active materials are spongy lead, Pb (for the negative plate), lead dioxide, PbO_2 (the for positive plate), and sulphuric acid. The active materials of the plates are supported in grids of hard lead-antimony alloy (Fig. 34.24 (i)). These are assembled in interchanging groups, closely spaced to give a low internal resistance, and often held apart by strips of wood or celluloid (Fig. 34.24 (ii)).

When the cell is discharging—giving a current—hydrogen ions drift to the positive plate, and SO_4^{2-} ions to the negative. As they give up their charges they attack the plates, and reduce the active materials of each to lead sulphate.

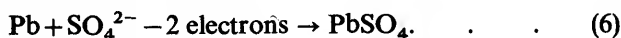


(i) Structure of plates

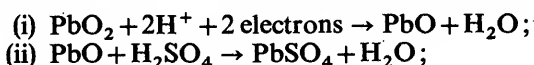
(ii) Assembly of cell

FIG. 34.24. Lead-acid accumulator.

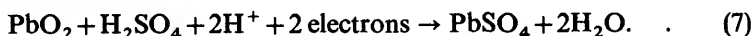
At the negative plate the reaction is



The chemical action at the positive plate is generally given as



whence, altogether



However, H_2SO_4 molecules do not exist in the solution—they are dissociated into 2H^+ and SO_4^{2-} ions. We may therefore write equation (7) as



The lead sulphate produced in these reactions is a soft form, which is chemically more active than the hard, insoluble lead sulphate familiar in the general chemistry of lead. In the discharging reactions water is formed and sulphuric acid consumed: the concentration of the acid, and therefore its specific gravity, fall.

Charging the Accumulator

When the cell is to be charged it is connected, in opposition, to a supply of greater e.m.f., via a rheostat and ammeter (Fig. 34.25). The

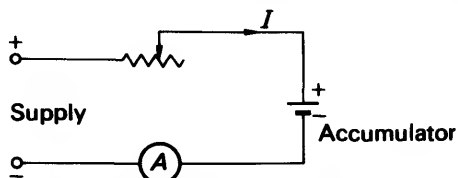
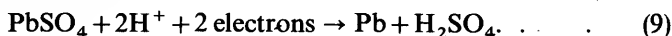


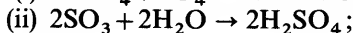
FIG. 34.25 Charging an accumulator.

supply forces a current I through the cell in the opposite direction to the discharging current, so that hydrogen ions are carried to the negative plate, and SO_4^{2-} ions to the positive. The chemical reactions are as follows.

At the negative plate:



At the positive plate:



altogether:



The active materials are converted back to lead and lead dioxide, water is consumed, and sulphuric acid is formed. The acid therefore becomes more concentrated during charge, and its specific gravity rises.

Properties and Care of the Lead Accumulator

The e.m.f. of a freshly charged lead accumulator is about 2.2 volts, and the specific gravity of the acid about 1.25. When the cell is being discharged its e.m.f. falls rapidly to about 2 volts, and then becomes steady (Fig. 34.26); but towards the end of the discharge the e.m.f. begins to fall again. When the terminal voltage load has dropped below about 1.9 volts, or the specific gravity of the acid below about 1.15, the cell should be recharged. If the cell is discharged too far, or left in a discharged condition, hard lead sulphate forms on its plates, and it becomes useless.

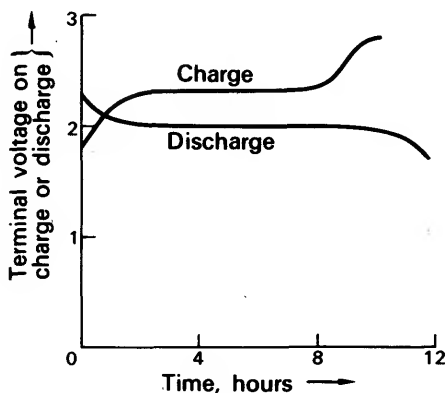


FIG. 34.26. Voltage-time curves of lead accumulator.

The internal resistance of a lead accumulator, like that of any other cell, depends on the area and spacing of its plates. It is much lower

than that of any primary cell, however, being usually of the order of $1/10$ to $1/100$ ohm. The amount of electricity which an accumulator can store is called its *capacity*. It is a vague quantity, but a particular accumulator may give, for example, 4 amperes for 20 hours before needing a recharge. The capacity of this accumulator would be 80 ampere-hours. (One ampere-hour = 3600 coulombs.) If the accumulator were discharged faster—at 8 amperes, say—then it would probably need recharging after rather less than 10 hours; and if it were discharged more slowly—say at 2 amperes—it might hold out for more than 40 hours. The capacity of an accumulator therefore depends on its rate of discharge; it is usually specified at the '10-hour' or '20-hour' rate. Discharging an accumulator faster than at about the 10-hour rate causes the active material to fall out of the plates.

Accumulators are usually charged at about the '8-hour' rate—say 5 amperes for the cell discussed above. The charging is continued until gas is bubbling freely off the plates. When the plates are gassing, the chemical reactions (9) and (10) have been completed, and the current through the cell is simply decomposing the water in it. Before the charge is started the vent-plugs in the cell-case must be removed to let the gases out; the gases are hydrogen and oxygen and naked lights near are dangerous. The water lost at the end of each charge must be made up by pouring in distilled water until the acid rises to the level marked on the case. If the specific gravity of the acid is then less than 1.25, the charging must be continued. Near the end of the charging, the back-e.m.f. of the cell rises sharply to about 2.6 volts (Fig. 34.26). It never gives a forward e.m.f. as great as this: as soon as it is put on discharge, its e.m.f. falls to about 2.2 volts.

Efficiency of Accumulator

The number of ampere-hours put into an accumulator on charge is greater than the number which can be got out of it without discharging it too far. The ratio

$$\frac{\text{ampere-hours on discharge}}{\text{ampere-hours on charge}}$$

is called the ampere-hour efficiency of the cell; its value is commonly about 90 per cent. However, to judge an accumulator by its ampere-hour efficiency is to flatter it; not only does it take in more ampere-hours on charge than it gives out on discharge, but it takes them in at a higher voltage. The electrical energy put into an accumulator on charge is the integral of current, e.m.f., and time:

$$W = \int I E dt.$$

For simplicity we may say

$$\begin{aligned} \text{energy put in} &= \text{quantity of electricity put in} \\ &\quad \times \text{average e.m.f. on charge.} \end{aligned}$$

If the quantity of electricity is measured in ampere-hours, the energy is in watt-hours instead of joules. Similarly, on discharge,

$$\text{energy given out} = \text{quantity of electricity given out} \times \text{average e.m.f. on discharge.}$$

The energy efficiency of the cell is

$$\begin{aligned} & \frac{\text{energy given out on discharge}}{\text{energy taken in on charge}} \\ &= \frac{\text{amp-hours} \times \text{average e.m.f. on discharge}}{\text{amp-hours} \times \text{average e.m.f. on charge}} \\ &= \text{amp-hour efficiency} \times \frac{\text{average e.m.f. on discharge}}{\text{average e.m.f. on charge}} \\ &\approx \text{amp-hour efficiency} \times \frac{2.0}{2.2} \end{aligned}$$

The energy efficiency is more often called the watt-hour efficiency of the cell; it is about 80 per cent.

The Nickel-iron Accumulator

The nickel-iron (NIFE) accumulator has active materials of nickel hydroxide (positive), iron (negative) and caustic potash solution. Its e.m.f. varies from 1.3 volt to 1.0 on discharge; it has a higher internal resistance than a lead accumulator of similar size; and it is less efficient. Its advantages are that it is more rugged, both mechanically and electrically. Very rapid charging and discharging do not harm it, nor do overdischarging and overcharging. Vibration does not make the active materials fall out of the plates, as it does with a lead cell. Nickel-iron accumulators are therefore used in electric trucks and at sea.

EXAMPLES

1. A copper refining cell consists of two parallel copper plate electrodes, 6 cm apart and 1 metre square, immersed in a copper sulphate solution of resistivity 1.2×10^{-2} ohm metre. Calculate the potential difference which must be established between the plates to provide a constant current to deposit 480 g of copper on the cathode in one hour (e.c.e. of copper = 3.29×10^{-7} kg C⁻¹).

From $m = zIt$, since 480 g = 0.48 kg,

$$I = \frac{m}{zt} = \frac{0.48}{3.29 \times 10^{-7} \times 3600} \text{ A} \quad \dots \dots \dots (i)$$

$$\begin{aligned} \text{The resistance of the cell, } R, &= \frac{\rho l}{A} \\ &= \frac{1.2 \times 10^{-2} \times 6 \times 10^{-2}}{1^2} \end{aligned}$$

$$\begin{aligned} \text{Hence, from (i), the p.d. } V = IR &= \frac{0.48 \times 1.2 \times 10^{-2} \times 6 \times 10^{-2}}{3.29 \times 10^{-7} \times 3600 \times 1^2} \\ &= 0.3 \text{ V (approx.).} \end{aligned}$$

2. State Faraday's laws of electrolysis and show that the ionic dissociation theory offers an explanation of them. Acidulated water is electrolyzed between platinum electrodes. Sketch a graph showing the relation between the strength of the current and the reading of a voltmeter connected to the electrodes. Comment on the nature of the graph.

Give a circuit diagram showing how you would charge a series battery of 12 lead accumulators, each of e.m.f. 2 volts and internal resistance $1/24$ ohm, from 240-volt d.c. mains, if the charging current is not to exceed 3 amp. What percentage of the energy taken from the mains would be wasted? (L.)

First part (see text). When the water is electrolyzed, no current flows until the p.d. is greater than about 1.7 volts, when the back-e.m.f. of the liberated product is overcome. After this, a straight-line graph is obtained between V and I .

Second part. A series resistance R is required, given by

$$I = 3 = \frac{240 - 12 \times 2}{R + \frac{12}{24}}$$

$$\therefore 3R + 1.5 = 216.$$

$$\therefore R = \frac{214.5}{3} = 71.5 \text{ ohms.}$$

Energy taken from mains = $EIt = 240 \times 3t = 720t$, where t is the time.

$$\begin{aligned} \text{Energy wasted} &= (I^2R + I^2r)t = (3^2 \times 71.5 + 3^2 \times 0.5)t \\ &= 648t. \end{aligned}$$

$$\therefore \text{percentage wasted} = \frac{648t}{720t} \times 100\% = 90\%.$$

3. State Faraday's laws of electrolysis. How would you verify the laws experimentally? Discuss briefly the phenomenon of polarization in electrolysis and how it is overcome in the Daniell cell.

Calculate a value for the e.m.f. of a Daniell cell from energy considerations, using the following data: 1 g of zinc dissolved in copper sulphate solution liberates 3327 J (e.c.e. of zinc = $0.000340 \text{ g C}^{-1}$). (L.)

The e.m.f. E of a cell can be defined as the energy per coulomb delivered by the cell (p. 797).

When 1 g of zinc is dissolved, number of coulombs flowing, $Q = \frac{1}{0.00034}$

Energy liberated, $W = 3327 \text{ J}$.

$$\begin{aligned} \therefore \text{energy liberated per coulomb} &= \frac{W}{Q} \\ &= 3327 \times 0.00034 \text{ joules/coulomb} \\ &= 1.13 \text{ volts.} \end{aligned}$$

EXERCISES 34

1. State Faraday's laws of electrolysis and show how they can be interpreted in terms of the ionic theory.

A voltmeter with large platinum plates contains dilute sulphuric acid in which 10^{-2} g of metallic copper has been dissolved. A constant current of 10^{-2} amp is then passed through the solution for 25 minutes. Calculate (a) the mass of copper deposited on the cathode, (b) the volume of oxygen liberated at the anode if the pressure is 750 mm Hg and the temperature is 20°C and (c) the additional time for which the same current must be passed before hydrogen is liberated at the cathode. (The chemical equivalents of copper and oxygen are 31.8 and 8

respectively; the Faraday is 95600 coulombs; 32 g of oxygen occupy 22.4 litres at S.T.P.) (O. & C.)

2. Describe the electrolytic processes which occur in a Daniell cell when its terminals are joined through a small resistance.

A steady current of 5 A is passed through a silver voltameter in series with a coil of wire of 10 ohms resistance immersed in 200 g of water. What will be the rise of temperature of the latter when 0.10 g of silver has been deposited? (Assume that the e.c.e. of silver = 1.118×10^{-6} kg C⁻¹; thermal capacity of the coil and vessel = 42 J K⁻¹.) (L.)

3. State Faraday's laws of electrolysis. Explain why it is necessary to have a potential greater than about 1.5 volts in order to maintain a large steady current through acidulated water.

In the electrolysis of water 83.7 cm³ of hydrogen were collected at a pressure of 68 cm of mercury at 25°C when a current of 0.5 A had been passed for 20 minutes. What is the electrochemical equivalent of copper in copper sulphate (CuSO₄)? (Atomic weight of copper = 63.57, atomic weight of hydrogen = 1.008, density of hydrogen at S.T.P. = 0.08987 kg m⁻³.) (L.)

4. What do you understand by (a) a *cathode*, and (b) an *ion*? Describe how the simple ionic theory accounts for the fact that the liberation of 1 g of hydrogen in the electrolysis of dilute sulphuric acid, using platinum electrodes, always requires the passage of a definite quantity of charge.

If the heat produced when 1 g of hydrogen burns to form water is 1.44×10^5 joules, calculate the minimum e.m.f. which must be applied in such electrolysis before a continuous current will flow. Justify and explain your calculation by reference to the principle of conservation of energy. (1 Faraday = 9.65×10^4 coulombs.)

Describe an experiment to determine the actual value of the back e.m.f. generated by a gas at the electrodes in such electrolysis. (C.)

5. State (a) Faraday's laws of electrolysis, (b) the main features of the theory of ionic dissociation, showing that the theory is in accordance with these laws.

A potential difference of 14 volts is applied to a water voltameter whose total thermal capacity is 2121 J K⁻¹. The temperature of the voltameter rises 1°C in the time taken to liberate 20 cm³ of hydrogen at S.T.P. Calculate the back e.m.f. of the voltameter and state the source of this e.m.f. (Assume the electrochemical equivalent of hydrogen to be 1.044×10^{-8} kg C⁻¹ and its density at S.T.P. to be 9×10^{-2} kg m⁻³.) (L.)

6. To answer this question you are asked to assume that no galvanometer of any kind is available and that you cannot take for granted that the same current flows continuously round a series circuit.

A circuit consists only of a straight horizontal length of wire and a copper voltmeter connected in series to a suitable d.c. supply. How would you show that a current is flowing through the wire as well as through the voltmeter?

How would you show that the direction of the current through the wire is the same as that of the current through the voltmeter?

Outline the measurements you would make to discover whether the strength of the current in the wire is the same as that of the current through the voltmeter.

Give an account of the actual physical processes by which a current passes through a metal wire. Explain briefly why the resistance normally increases as the temperature rises. (O.)

7. Explain the general nature of the chemical changes that take place in a lead accumulator during charging and discharging.

A battery of accumulators, of e.m.f. 50 volts and internal resistance 2 ohms, is charged on a 100-volt direct-current mains. What series resistance will be required to give a charging current of 2 A? If the price of electrical energy is 1d. per kilowatt-hour, what will it cost to charge the battery for 8 hours, and what percentage of the energy supplied will be wasted in the form of heat? (C.)

8. State the laws of electrolysis and give a concise account of an elementary theory of electrolysis which is consistent with the laws.

If an electric current passes through a copper voltameter and a water voltameter in series, calculate the volume of hydrogen which will be liberated in the latter, at 25°C and 78 cm of mercury pressure, whilst 0.05 g of copper is deposited in the former. (Take e.c.e. of hydrogen as 1.04×10^{-8} kg C⁻¹, e.c.e. of copper as 3.3×10^{-7} kg C⁻¹, density of hydrogen as 9×10^{-2} kg m⁻³ at S.T.P.) (L.)

9. State Faraday's laws of electrolysis and describe experiments to verify them.

A difference of potential of 60 volts is maintained between two electrodes 12 cm apart in a solution of common salt. How long will it take a chlorine ion to travel 3 cm in the solution? (The mobility of chlorine ions may be taken as 0.00053 cm s⁻¹ per V cm⁻¹.) (L.)

10. Explain what happens when an e.m.f. is applied to platinum electrodes immersed in dilute sulphuric acid. What is the relation between the e.m.f. and the current in such a cell?

If the electrochemical equivalent of hydrogen is 1.04×10^{-8} kg coulomb⁻¹, and if 1 g of hydrogen on burning to form water liberates 147 000 joules, calculate the back-e.m.f. produced in a water voltameter when it is connected to a 2-volt accumulator. (C.)

chapter thirty-five

Magnetic Field and Force on Currents

NATURAL magnets were known some thousands of years ago, and in the eleventh century A.D. the Chinese invented the magnetic compass. This consisted of a magnet, floating on a buoyant support in a dish of water. The respective ends of the magnet, where iron filings are attracted most, are called the north and south poles.

In the thirteenth century the properties of magnets were studied by Peter Peregrinus. He showed that

like poles repel and unlike poles attract

His work was forgotten, however, and his results were rediscovered in the sixteenth century by Dr. Gilbert, who is famous for his researches in magnetism and electrostatics.

Ferromagnetism

About 1823 Sturgeon placed an iron core into a coil carrying a current, and found that the magnetic effect of the current was increased enormously. On switching off the current the iron lost nearly all its magnetism. Iron, which can be magnetized strongly, is called a *ferromagnetic* material. Steel, made by adding a small percentage of carbon to iron, is also ferromagnetic. It retains its magnetism, however, after removal from a current-carrying coil, and is more difficult to magnetize than iron. Nickel and cobalt are the only other ferromagnetic elements in addition to iron, and are widely used for modern magnetic apparatus. A modern alloy for permanent magnets, called *alnico*, has the composition 54 per cent iron, 18 per cent nickel, 12 per cent cobalt, 6 per cent copper, 10 per cent aluminium. It retains its magnetism extremely well, and, by analogy with steel, is therefore said to be magnetically very hard. Alloys which are very easily magnetized, but do not retain their magnetism, are said to be magnetically soft. An example is *mumetal*, which contains 76 per cent nickel, 17 per cent iron, 5 per cent copper, 2 per cent chromium.

Magnetic Fields

The region round a magnet, where a magnetic force is experienced, is called a *magnetic field*. The appearance of a magnetic field is quickly obtained by iron filings, and accurately plotted with a small compass, as the reader knows. The *direction* of a magnetic field is taken as the direction which a north pole would move if placed in the field.

Fig. 35.1 shows a few typical fields. The field round a bar-magnet is 'non-uniform', that is, its strength and direction vary from place to

place (Fig. 35.1 (i)). The earth's field locally, however, is uniform (Fig. 35.1 (ii)). A bar of soft iron placed north-south becomes magnetized by

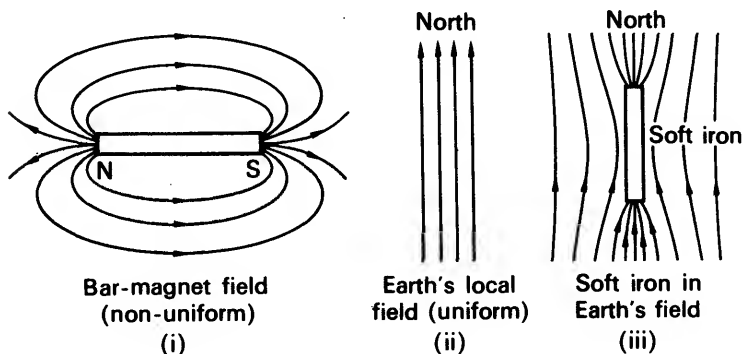


FIG. 35.1. Magnetic fields.

induction by the earth's field, and the lines of force become concentrated in the soft iron (Fig. 35.1 (iii)). The tangent to a line of force at a point gives the direction of the magnetic field at that point.

Oersted's Discovery

The magnetic effect of the electric current was discovered by Oersted in 1820. Like many others, Oersted suspected a relationship between electricity and magnetism, and was deliberately looking for it. In the course of his experiments, he happened to lead a wire carrying a current

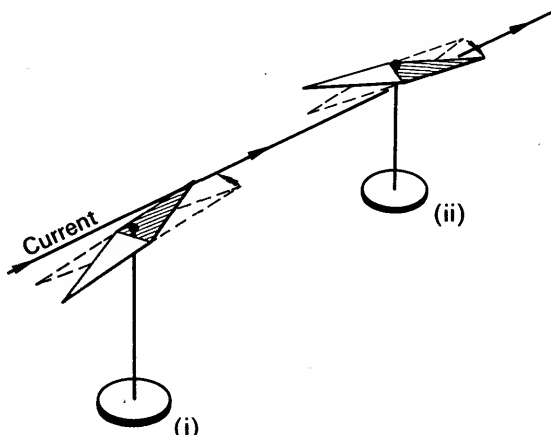


FIG. 35.2. Deflection of compass-needle by electric current.

over, but parallel to, a compass-needle, as shown in Fig. 35.2(i); the needle was deflected. Oersted then found that if the wire was led under

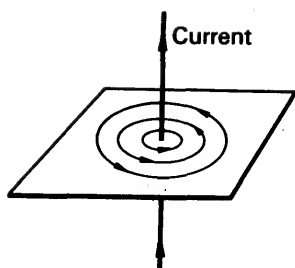


FIG. 35.3. Magnetic field of long straight conductor.

the needle, it was deflected in the opposite sense (Fig. 35.2(ii)). From these observations he concluded that 'the electric conflict performs gyrations'. What he meant by this we can see by plotting the lines of force of a long vertical wire, as shown in Fig. 35.3. To get a clear result a strong current is needed, and we must work close to the wire, so that the effect of the earth's field is negligible. It is then seen that the lines of force are circles, concentric with the wire.

Directions of Current and Field; Corkscrew Rule

The relationship between the direction of the lines of force and of the current is expressed in Maxwell's corkscrew rule: if we imagine ourselves driving a corkscrew in the direction of the current, then the direction of rotation of the corkscrew is the direction of the lines of force. Fig. 35.4 illustrates this rule, the small, heavy circle representing the wire, and the large light one a line of force. At (i) the current is flowing into the paper; its direction is indicated by a cross, which stands for the tail of a retreating arrow. At (ii) the current is flowing out of the paper; the dot in the centre of the wire stands for the point of an approaching arrow.

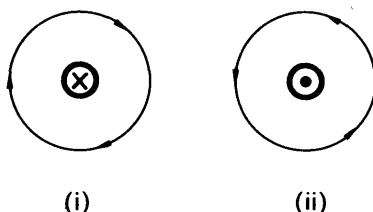


FIG. 35.4. Illustrating corkscrew rule.

If we plot the magnetic field of a circular coil carrying a current, we get the result shown in Fig. 35.5. Near the circumference of the coil, the lines of force are closed loops, which are not circular, but whose

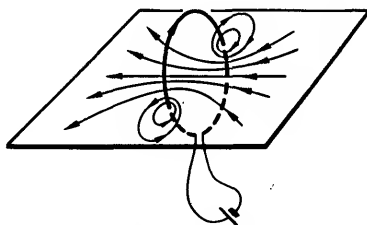


FIG. 35.5. Magnetic field of narrow coil.

directions are still given by the corkscrew rule, as in Fig. 35.5. Near the centre of the coil, the lines are almost straight and parallel. Their

direction here is again given by the corkscrew rule, but the current and the lines of force are interchanged: if we turn the screw in the direction of the current, then it travels in the direction of the lines.

The Solenoid

The same is true of the magnetic field of a long cylindrical coil, shown in Fig. 35.6. Such a coil is called a solenoid; it has a field similar to that of a bar-magnet, whose poles are indicated in the figure. If an iron or steel core were put into the coil, it would become magnetized, with the polarity shown.

If the terminals of a battery are joined by a wire which is simply

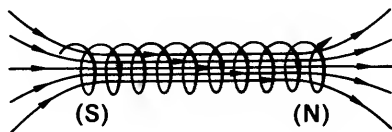


FIG. 35.6. Magnetic field of solenoid.

doubled back on itself, as in Fig. 35.7, there is no magnetic field at all: each element of the outward run, such as AB, in effect cancels the field of the corresponding element of the inward run, CD. But as soon as the wire is opened out into a loop, its magnetic field appears (Fig. 35.8).

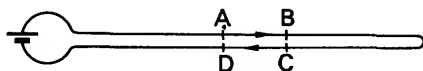


FIG. 35.7. A double-back current has no magnetic field.

Within the loop, the field is strong, because all the elements of the loop give magnetic fields in the same sense, as we can see by applying the corkscrew rule to each side of the square ABCD. Outside the loop, for example at the point P, corresponding elements of the loop give

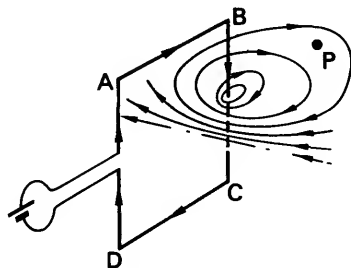


FIG. 35.8. An open loop of current has magnetic field.

opposing fields (for example, DA opposes BC); but these elements are at different distances from P (DA is farther away than BC). Thus

there is a resultant field at P, but it is weak compared with the field inside the loop. A magnetic field can thus be set up either by wires carrying a current, or by the use of permanent magnets.

Force on Conductor. Fleming's Rule

When a conductor carrying a current is placed in a magnetic field due to some source other than itself, it experiences a mechanical force.

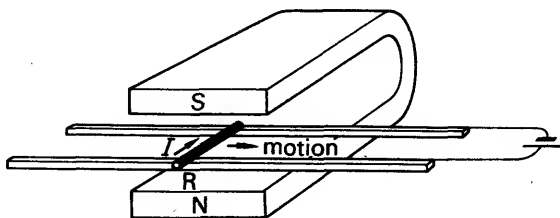


FIG. 35.9. Force on current in magnetic field.

To demonstrate this, a short brass rod R is connected across a pair of brass rails, as shown in Fig. 35.9. A horseshoe magnet is placed so that the rod lies in the field between its poles. When we pass a current through the rod, from an accumulator, the rod rolls along the rails. The relative directions of the current, the applied field, and the motion are shown in Fig. 35.10; they are the same as those of the middle finger, the fore-finger, and the thumb of the left hand when held all at right angles to one another. If we place the horseshoe magnet so that its field lies along the rod carrying the current, then the rod experiences no force.

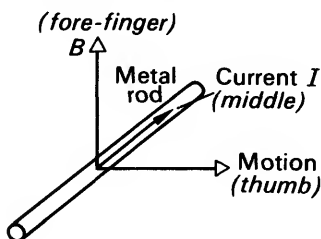


FIG. 35.10. Left-hand rule.

Experiments like this were first made by Ampère in 1820. As a result of them, he concluded that the force on a conductor is always *at right angles to the plane which contains both the conductor and the direction of the field in which it is placed*. He also showed that, if the conductor makes an angle α with the field, the force on it is proportional to $\sin \alpha$, so that the maximum force is exerted when the conductor is perpendicular to the field.

Dependence of Force on Physical Factors

Since the magnitude of the force on a current-carrying conductor is given by

$$F \propto \sin \alpha, \quad (1)$$

where α is the angle between the conductor and the field, it follows that F is zero when the conductor is parallel to the field direction. This defines the direction of the magnetic field. To find which way it points,

we can apply Fleming's rule to the case when the conductor is placed at right angles to the field. The sense of the field then corresponds to the direction of the forefinger.

Variation of F with I and l

To investigate how the magnitude of the force F depends on the current I and the length l of the conductor, we may use the apparatus of Fig. 35.11.

Here the conductor AC is situated in the field of a solenoid S. The current flows into, and out of, the wire via the pivot points X and Y. The scale pan T is placed at the same distance from the pivot as the straight wire AC, which is perpendicular to the axis of the coil. The

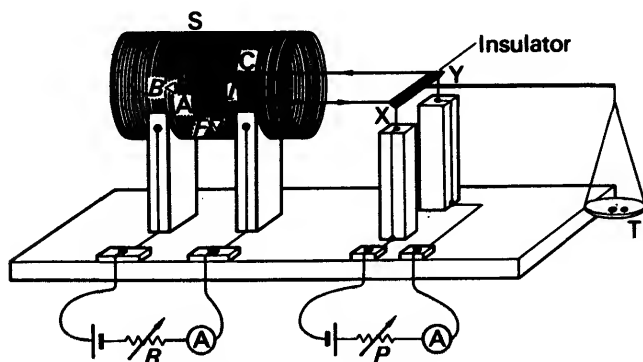


FIG. 35.11. Experiment to show how F varies with I and B .

frame is first balanced with no current flowing in AC. A current is then passed, and the extra weight needed to restore the frame to a horizontal position is equal to the force on the wire AC. By varying the current in AC with the rheostat P , it may be shown that:

$$F \propto I \quad . \quad . \quad . \quad . \quad . \quad (2)$$

If different frames are used so that the length, l , of AC is changed, it can be shown that

$$F \propto l \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Effect of B

The magnetic field due to the solenoid will depend on the current flowing in it. If this current is varied by adjusting the rheostat R , it can be shown that the larger the current in the solenoid, S , the larger is the force F . It is reasonable to suppose that a larger current in S produces a stronger magnetic field. Thus the force F increases if the magnetic field strength is increased. The magnetic field is represented by a vector quantity which is given the symbol B . This is called the *magnetic induction* or *flux density* in the field. We assume that:

$$F \propto B \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Magnitude of F

From the results expressed in equations (1) to (4), we obtain

$$F \propto BIl \sin \alpha,$$

or

$$F = \kappa BIl \sin \alpha \quad (5)$$

where κ is a constant.

In the SI system of units, the unit of B is the tesla (T). $1 \text{ T} = 1 \text{ weber metre}^{-2} (\text{Wb m}^{-2})$. One tesla may be defined as the flux density in a field when the force on a conductor 1 metre long, placed perpendicular

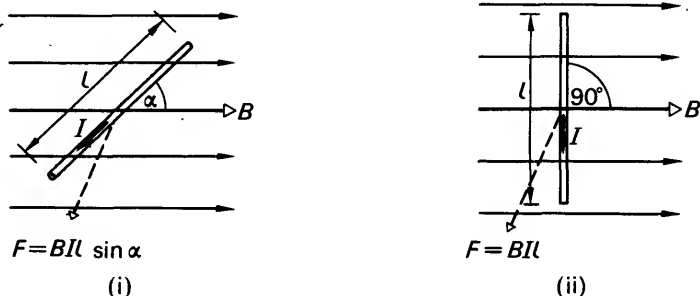


FIG. 35.12. Magnitude of F .

to the field and carrying a current of 1 ampere, is 1 newton. Substituting $F = 1$, $B = 1$, $l = 1$ and $\sin \alpha = \sin 90^\circ = 1$ in (5), then $\kappa = 1$. Thus in Fig. 35.12 (i), with the above units,

$$F = BIl \sin \alpha \quad (5)$$

When the whole length of the conductor is perpendicular to the field B , Fig. 35.12 (ii), then, since $\alpha = 90^\circ$ in this case,

$$F = BIl \quad (6)$$

It may be noted that the apparatus of Fig. 35.11 can be used to determine the flux density B of the field in the solenoid. In this case, $\alpha = 90^\circ$ and $\sin \alpha = 1$, so that measurement of F , I and l enables B to be found from (6).

EXAMPLE

A wire carrying a current of 10 A and 2 metres in length is placed in a field of flux density $0.15 \text{ T} (\text{Wb m}^{-2})$. What is the force on the wire if it is placed (a) at right angles to the field, (b) at 45° to the field, (c) along the field.

From (5)

$$F = BIl \sin \alpha$$

$$(a) \quad \therefore F = 0.15 \times 10 \times 2 \times \sin 90^\circ = 3 \text{ newtons.}$$

$$(b) \quad \therefore F = 0.15 \times 10 \times 2 \times \sin 45^\circ = 2.12 \text{ newtons.}$$

$$(c) \quad \therefore F = 0, \text{ since } \sin 0^\circ = 0.$$

Force on Moving Charges

It was explained in Chapter 32 that an electric current in a wire can be regarded as a drift of electrons in the wire, superimposed on their random thermal motions. If the electrons in the wire drift with average velocity v , and the wire lies at right angles to the field, then the force on each electron is given by

$$F = Bev \quad (1)$$

If B is in tesla (T) or weber metre⁻² (Wb m⁻²), e is in coulomb (C) and v in metre second⁻¹ (m s⁻¹), then F will be in newtons (N) (Fig. 35.13(i)).

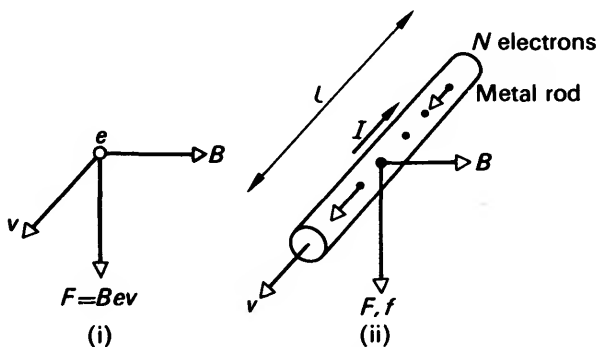


FIG. 35.13. Force on moving electron in magnetic field (v at right angles to page).

The proof of equation (1) is easily obtained. If there are N free electrons in a wire of length l , their total charge is Ne , Fig. 35.13 (ii). If they have a drift velocity v , the time which any one of them takes to travel the distance l is

$$t = \frac{l}{v}.$$

In this time, therefore, the N electrons are swept out of the wire by the current and are replaced by another N electrons. The rate at which charge flows along the wire, which is the current through it, is therefore

$$I = \frac{Ne}{t} = \frac{Nev}{l}.$$

If the wire is at right angles to a magnetic field B , the force on it is

$$f = BIl = BNev.$$

Therefore the force on a single electron is

$$F = \frac{f}{N} = Bev.$$

An electron moving across a magnetic field experiences a force whether it is in a wire or not—for example, it may be one of a beam of electrons in a vacuum tube. Because of this force, a magnetic field can be used to focus or deflect an electron beam, instead of an electrostatic field as on p. 1002. Magnetic deflection and focusing are common in cathode ray tubes used for television.

The Moving-coil Galvanometer

All except the most accurate of current measurements are made today with the moving-coil galvanometer. In this instrument a coil of fine insulated copper wire, $ABDF$ in Fig. 35.14 (i), hangs in a strong magnetic field. The field is set up between soft iron pole-pieces, NS , attached to a powerful permanent magnet.

The pole-pieces are curved to form parts of a cylinder coaxial with the suspension of the coil. And between them lies a cylindrical core of soft iron, C ; it is supported on a brass pin, T in Fig. 35.14(iii), which is placed so that it does not foul the coil. As this figure shows, the magnetic field B is radial to the core and pole-pieces, over the region in which the coil can swing.

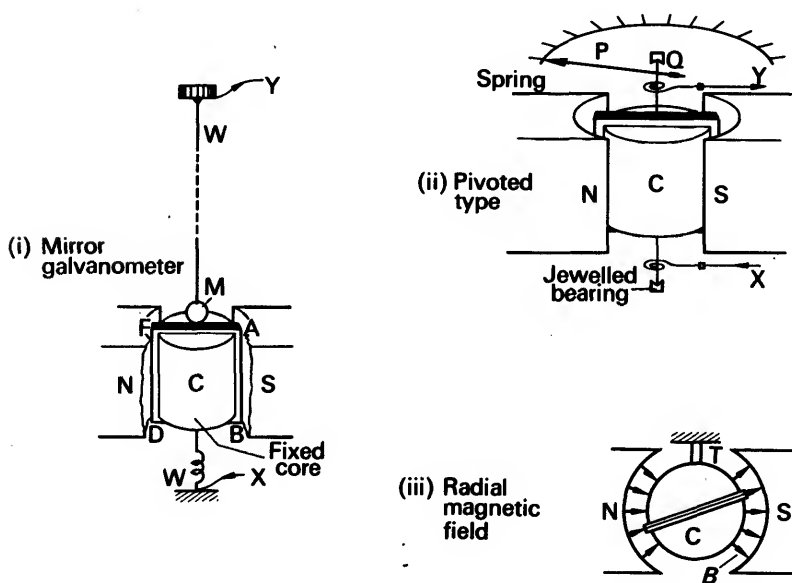


FIG. 35.14. Moving-coil galvanometers.

In the more sensitive instruments, the coil is suspended on a phosphor-bronze wire, WW , which is kept taut. The current is led into and out of the coil through the suspension, at X and Y , and the deflection of the coil is shown by a mirror, M . This is known as a *mirror galvanometer*. More robust but less sensitive forms of the galvanometer have hair-springs and jewelled bearings, instead of the phosphor-bronze suspension (Fig. 35.14 (ii)). The coil is wound on a rigid but light former, of bamboo or aluminium, which also carries the pivots. The pivots are insulated from the former if it is aluminium, and the current is led in and out through the springs. The framework, which

carries the springs and jewels, is made from brass or aluminium—if it were steel it would affect the magnetic field. An aluminium pointer, P, shows the deflection of the coil; it is balanced by a counterweight, Q.

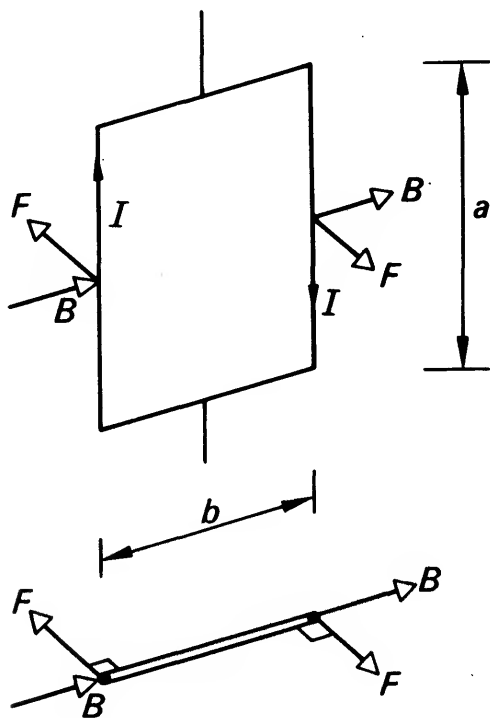


FIG. 35.15. Couple on coil in radial field.

Theory of Moving-coil Instrument

Fig. 35.15 shows a current I flowing through the coil of a moving-coil instrument. The magnetic field exerts forces F on the vertical limbs of each turn of the coil. These forces tend to rotate the coil about the suspension. Since the magnetic field B is at right angles to the vertical limbs, the forces are given by

$$F = BIl,$$

where l is the height of the coil; F acts at right angles to its plane because the magnetic field is radial. Therefore, if b is the width of the coil, the couple which the forces F exert on each turn is:

$$C = Fb = BIlb = BIA,$$

If the coil has N turns, then the total couple which the current exerts on it is

$$C' = NC = BIAN \quad (1)$$

The coil will turn until the restoring couple due to the twist in the suspension is equal to C' . This couple is proportional to the twist θ , which is also the deflection of the coil; thus

$$C' = k\theta,$$

where k is a constant of the suspension. Hence

$$BIAN = k\theta$$

or
$$I = \frac{k}{BAN} \theta. \quad (2)$$

Equation (2) shows that the deflection is proportional to the current. The pointer type of instrument (Fig. 35.15 (ii)) usually has a scale calibrated directly in milliamperes or microamperes. Full-scale reading on such an instrument corresponds to a deflection θ of 90° to 120° ; it may represent a current of 50 microamperes to 15 milliamperes, according to the strength of the hair-springs, the geometry of the coil, and the strength of the magnetic field. The less sensitive models are more accurate, because their pivots and springs are more robust, and therefore are less affected by dust, vibration, and hard use. Models known as 'first grade' (FG) have an error not greater than 1 per cent of full-scale deflection. This error is constant over the scale, so that the inaccuracy may be 10 per cent when the reading is one-tenth of full scale. Readings less than about half full scale must therefore be avoided in accurate work. The best moving-coil instruments have an error of about 0.2 per cent of full-scale deflection.

When a galvanometer is of the suspended-coil type (Fig. 35.15 (i)), its sensitivity is generally expressed in terms of the displacement of the spot of light reflected from the mirror on to the scale. At a scale distance of 1 metre a moderately sensitive instrument of this type will give a deflection of 100 cm per microampere.

All forms of moving-coil galvanometer have one disadvantage: they are easily damaged by overload. A current much greater than that which the instrument is intended to measure will burn out its hair-springs or suspension.

Couple on a Coil in a Uniform Field

So far we have considered a coil in a radial magnetic field. Now let us consider one in a uniform field. Fig. 35.16 shows a rectangular coil of one turn, whose plane makes an angle α with a uniform magnetic field B . If it carries a current I amperes, the forces F_1 on its vertical limbs are given by

$$F_1 = BIl = BIa,$$

where a is the height of the coil. And the forces F_2 on its horizontal limbs are given by

$$F_2 = Blb \sin \alpha,$$

where b is the width of the coil. The forces F_2 merely compress the coil, and are resisted by its rigidity. The forces F_1 set up a couple, whose moment is, from the lower figure,

$$\begin{aligned} C &= F_1 \times GN \\ &= F_1 b \cos \alpha. \end{aligned}$$

Hence

$$\begin{aligned} C &= BIab \cos \alpha \\ &= BIA \cos \alpha, \end{aligned} \quad (1)$$

where A is the area of the coil. If θ is the angle between the field and the normal to the plane of the coil, then $\theta = 90^\circ - \alpha$, and

$$C = BIA \sin \theta. \quad (2)$$

With a coil of N turns,

$$C = BNIA \quad (3)$$

Magnetic Moment

The couple on the coil thus depends on the magnitude of the flux density B , the current I in it and the Area A . We can write equation (3) as

$$C = mB \sin \theta \quad (1)$$

where $m = NIA$. m is a property of the coil and the current in it and is called the *magnetic moment* of the coil. (Fig. 35.17 (i)). In general, the magnetic moment of a coil is defined as the couple exerted on it when it is placed with its plane parallel to a field of unit induction. In this case $B = 1$, $\sin \theta = \sin 90^\circ = 1$ and hence $C = m$ (Fig. 35.17 (ii)). It should be noted that the magnetic moment of a coil can be calculated from

$$m = NIA \quad (2)$$

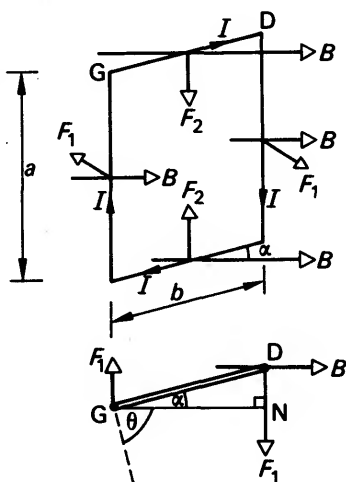


FIG. 35.16. Couple on coil in uniform field.

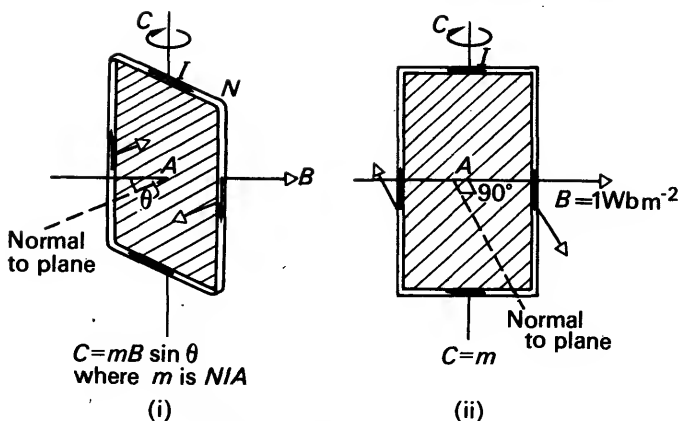


FIG. 35.17. Magnetic moment and couple.

whatever the shape of the coil. From this expression the unit of m is ampere metre² (A m²).

EXAMPLE

What couple is needed to hold a small single-turn coil of area 5 cm² in equilibrium, when it carries a current of 10 A and is placed with its axis at right angles to a field of flux density 0.15 T? What couple would be required if the coil had 20 turns?

$$\text{Area } A = 5 \text{ cm}^2 = 5 \times 10^{-4} \text{ m}^2.$$

$$\therefore \text{Magnetic moment of single-turn coil} = m = IA$$

$$= 10 \times (5 \times 10^{-5}) \text{ A m}^2.$$

$$\therefore C = mB \sin \theta$$

$$= (10 \times 5 \times 10^{-5}) \times 0.15 \times \sin 90^\circ$$

$$= 0.75 \times 10^{-4} \text{ newton metre (N m)}.$$

If the coil has 20 turns, each will have the same magnetic moment. Hence

$$\text{total couple} = 20 \times 0.75 \times 10^{-4}$$

$$= 1.5 \times 10^{-3} \text{ N m}.$$

Vibration Magnetometer

Magnetism is due to circulating and spinning electrons inside atoms (see p. 948). The moving charges are equivalent to electric currents. Consequently, like a current-carrying coil, permanent magnets also have a couple acting on them when they are placed with their axis at an angle to a magnetic field. Like the coil, they turn and settle in equilibrium with their axis along the field direction.

Consider a magnet suspended from a torsionless silk thread in a draught-shield. If it is gently disturbed, for example by momentarily bringing a piece of iron towards it, the magnet *vibrates* about the

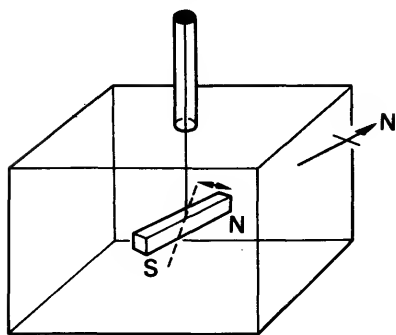


FIG. 35.18. Oscillation of bar-magnet.

magnetic meridian (Fig. 35.18).

The oscillation occurs because, when the magnet is deflected from the meridian, a couple acts on it, tending to bring it back into the meridian. On p. 885 it was shown that the couple had a moment $mB \sin \theta$, where m is the moment of the magnet, B is the field intensity, and θ is the angle between the magnet and the field. If the angular displacement, θ , is small—say not more than 10° —then, to a close approximation,

$\sin \theta = \theta$ in radians. The restoring couple C is then given by

$$C = mB\theta. \quad (1)$$

Thus the restoring couple is proportional to the angular displacement θ , and the magnet therefore makes simple harmonic vibrations (see p. 89).

Period of Vibration

The period of vibration, T , depends on the restoring couple per unit deflection of the magnet; from (1),

$$\frac{C}{\theta} = mB.$$

The period also depends on the moment of inertia K of the magnet, which is determined by its mass, size, and shape. Now from equation (1), p. 89,

$$T = 2\pi \sqrt{\frac{K}{\text{restoring couple per unit deflection}}}$$

$$\therefore T = 2\pi \sqrt{\frac{K}{mB}} \quad (1)$$

The number of vibrations made per second, n , or the frequency of the vibration, is given by

$$n = \frac{1}{T} = \frac{1}{2\pi} \sqrt{\frac{mB}{K}},$$

$$\therefore n^2 = \frac{mB}{4\pi^2 K},$$

or

$$B \propto n^2. \quad (2)$$

This relationship shows that a vibrating magnet may be used to compare magnetic flux densities, for example, to explore the earth's field over the laboratory. When used in this way it is called a *vibration magnetometer*.

It should be noted that the relations (1) and (2) are true only when the fibre suspending the magnet is torsionless, that is, it exerts a negligible restoring couple when it is twisted. The restoring couple on the magnet is then simply that due to the magnetic field ($mB \sin \theta$), as already assumed. A common material used for the suspension is unspun silk. The magnet must be small so that the whole of it is in the uniform field, and it must be protected from draughts, which would upset the oscillation.

Comparison of Flux Densities

By means of a vibration magnetometer, the values of B in two magnetic fields can easily be compared (see also p. 892). The magnetometer is set up in one field of magnitude B_1 say and the time period T_1 of small swings is measured. The new time period T_2 is then measured in the other field of magnitude B_2 . Then, from previous,

$$T_1 = 2\pi \sqrt{\frac{K}{mB_1}} \quad \text{and} \quad T_2 = 2\pi \sqrt{\frac{K}{mB_2}}$$

Dividing the two equations,

$$\therefore \frac{T_1}{T_2} = \sqrt{\frac{B_2}{B_1}} \quad \text{or} \quad \frac{B_1}{B_2} = \frac{T_2^2}{T_1^2}$$

Hence the flux densities in the two fields may be compared.

Comparison of Moments

The method of vibration can also be used to compare the magnetic moments of two magnets. The magnets are rigidly fixed together, with similar poles adjacent, and allowed to vibrate in the earth's field (Fig. 35.19 (i)). If m_1, m_2, K_1, K_2 , are their magnetic moments and moments

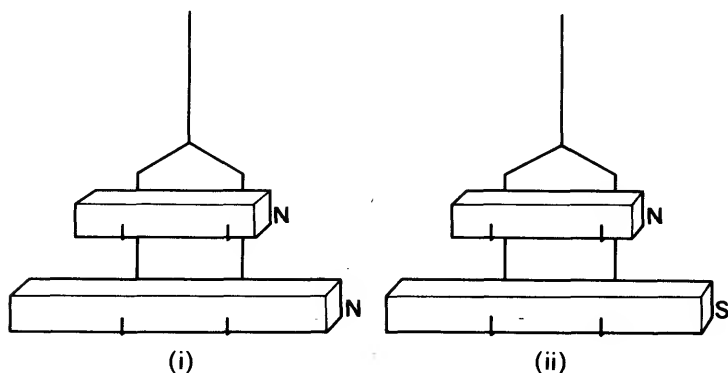


FIG. 35.19. Comparison of moments by vibration.

of inertia respectively, and if B_H is the earth's horizontal field, then the period of the magnets is

$$T_1 = 2\pi \sqrt{\frac{K_1 + K_2}{(m_1 + m_2)B_H}}$$

The magnets are now fixed together with opposite poles adjacent, and again allowed to vibrate. Fig. 35.19 (ii). Their period is now

$$T_2 = 2\pi \sqrt{\frac{K_1 + K_2}{(m_1 - m_2)B_H}}$$

Therefore

$$\frac{T_2}{T_1} = \sqrt{\frac{m_1 + m_2}{m_1 - m_2}},$$

whence

$$\frac{m_1}{m_2} = \frac{T_2^2 + T_1^2}{T_2^2 - T_1^2} \quad (1)$$

The Wattmeter

The wattmeter is an instrument for measuring electrical power. In construction and appearance it resembles a moving-coil voltmeter or ammeter, but it has no permanent magnet. Instead it has two fixed coils, FF in Fig. 35.20; these set up the magnetic field in which the suspended coil, M, moves. When the instrument is in use, the coils FF are connected in series with the device X whose power consumption is

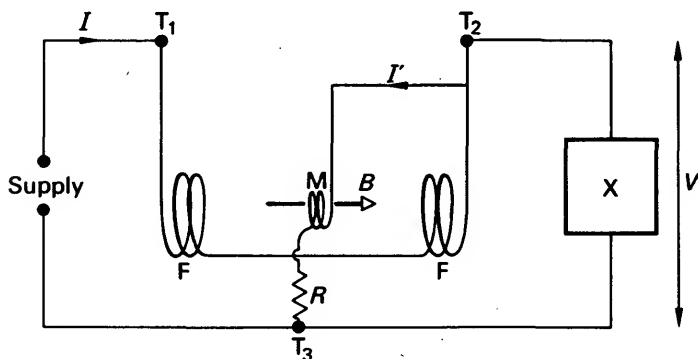


FIG. 35.20. Principle of wattmeter.

to be measured. The magnetic field B , set up by FF, is then proportional to the current I drawn by X:

$$B \propto I.$$

The moving coil M is connected across the device X. In series with M is a high resistance R, similar to the multiplier of a voltmeter; M is, indeed, often called the volt-coil. The current I' through the volt-coil

is small compared with the main current I , and is proportional to the potential difference V across the device X :

$$I' \propto V.$$

The couple acting on the moving coil is proportional to the current through it, and to the magnetic field in which it is placed:

$$C \propto BI'.$$

Consequently

$$C \propto IV.$$

That is to say, the couple on the coil is proportional to the product of the current through the device X , and the voltage across it. The couple is therefore proportional to the power consumed by X , and the power can be measured by the deflection of the coil.

The diagram shows that, because the volt-coil draws current, the current through the fixed coils is a little greater than the current through X . As a rule, the error arising from this is negligible; if not, it can be allowed for as when a voltmeter and ammeter are used separately.

Hall Effect

In 1879, Hall found that an e.m.f. is set up *transversely* or *across* a current-carrying conductor when a perpendicular magnetic field is applied. This is called the *Hall effect*.

To explain the Hall effect, consider a slab of metal carrying a current (Fig. 35.21). The flow of electrons is in the opposite direction to the conventional current. If the metal is placed in a magnetic field B at

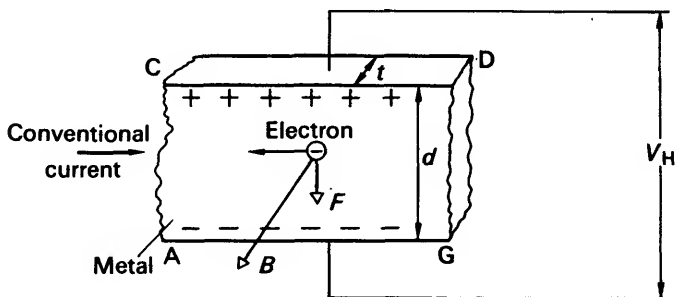


FIG. 35.21. Hall voltage.

right angles to the face AGDC of the slab and directed out of the plane of the paper, a force Bev then acts on each electron in the direction from CD to AG. Thus electrons accumulate along the side AG of the metal, which will make AG negatively charged and lower its potential with respect to CD. Thus a potential difference or e.m.f. opposes the electron flow. The flow ceases when the e.m.f. reaches a particular value V_H called the *Hall voltage*, which may be measured by using a high impedance voltmeter as shown in Fig. 35.22.

Magnitude of Hall Voltage

Suppose V_H is the magnitude of the Hall voltage and d is the width of the slab. Then the electric field intensity E set up across the slab is numerically equal to the potential gradient and hence $E = V_H/d$. Hence the force on each electron $= Ee = V_H e/d$.

This force, which is directed upwards from AG to CD, is equal to the force produced by the magnetic field when the electrons are in equilibrium.

$$\therefore Ee = Bev$$

$$\therefore \frac{V_H e}{d} = Bev$$

$$\therefore V_H = Bvd \quad \dots \dots \dots (1)$$

From p. 787, the drift velocity of the electrons is given by

$$I = NevA, \quad \dots \dots \dots (2)$$

where N is the number of electrons per unit volume and A is the area of cross section of the conductor. In this case $A = td$ where t is the thickness. Hence, from (2),

$$v = \frac{I}{Net}$$

Substituting in (1),

$$\therefore V_H = \frac{BI}{Net} \quad \dots \dots \dots (3)$$

We now take some typical values for copper to see the order of magnitude of V_H . Suppose $B = 1$ T, a field obtained by using a large laboratory electro-magnet. For copper, $N \approx 10^{29}$ electrons per metre³, and the charge on the electron is 1.6×10^{-19} coulomb. Suppose the specimen carries a current of 10 A and that its thickness is about 1 mm or 10^{-3} m. Then

$$V_H = \frac{1 \times 10}{10^{29} \times 1.6 \times 10^{-19} \times 10^{-3}} = 0.6 \mu\text{V (approx.)}$$

This e.m.f. is very small and would be difficult to measure. The importance of the Hall effect becomes apparent when semiconductors are used, as we now see.

Hall Effect in Semiconductors

In semiconductors, the charge carriers which produce a current when they move may be positively or negatively charged (see p. 1024). The Hall effect helps us to find the sign of the charge carried. In Fig. 35.21, p. 890, suppose that electrons were not responsible for carrying the current, and that the current was due to the movement of positive charges in the *same* direction as the conventional current. The magnetic force on these charges would also be downwards, in the same direction as if the current were carried by electrons. This is because the sign *and* the direction of movement of the charge carriers have both been reversed. Thus AB would now become positively charged, and the polarity of the Hall voltage would be reversed. Experimental investigation of the polarity of the Hall voltage hence tells us whether the current is predominantly due to the drift of positive charges or to the drift of negative

charges. In this way it was shown that the current in a metal such as copper is due to movement of negative charges, but that in impure semiconductors such as germanium or silicon, the current may be predominantly due to movement of either negative or positive charges (p. 945).

The magnitude of the Hall voltage V_H in metals was shown on p. to be very small. In semiconductors it is much larger because the number N of charge carriers per metre³ is much *less* than in a metal and $V_H = BI/Net$. Suppose that N is about 10^{25} per metre³ in a semiconductor, and $B = 1 \text{ T (Wb m}^{-2}\text{)}$, $t = 10^{-3} \text{ m}$, $e = 1.6 \times 10^{-19} \text{ C}$, as on p. 891. Then

$$V_H = \frac{1 \times 10}{10^{25} \times 1.6 \times 10^{-19} \times 10^{-3}} = 6 \times 10^{-3} \text{ V (approx.)} = 6 \text{ mV.}$$

The Hall voltage is thus much more measurable in semiconductors than in metals.

Use of Hall Effect

An instrument called a *Hall probe* may now be used to measure the flux density B of a magnetic field. A simple Hall probe is shown in Fig. 35.22. Here a wafer of semiconductor has two contacts on opposite

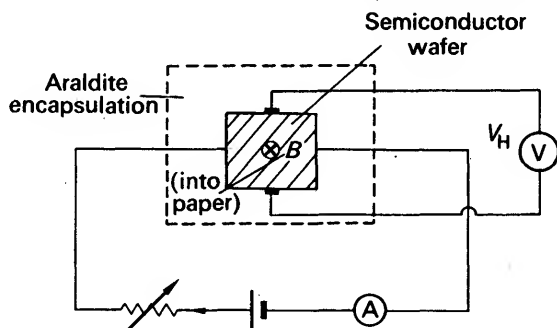


FIG. 35.22. B by Hall voltage.

sides which are connected to a high impedance voltmeter, V . A current, generally less than one ampere, is passed through the semiconductor and is measured on the ammeter, A . The 'araldite' encapsulation prevents the wires from being detached from the wafer. Now, from (3) on p. 891.

$$V_H = \frac{BI}{Net}$$

$$\therefore B = \frac{V_H Net}{I}$$

Now Net is a constant for the given semiconductor, which can be determined previously. Thus from the measurement of V_H and I , B can be found.

EXERCISES 35

1. With the aid of clear labelled diagrams describe the structure of a suspended moving-coil galvanometer and explain its mode of action. Derive an expression for its current sensitivity.

A certain galvanometer when shunted with a resistance 0.0500 ohm gives a full-scale deflection for 2 A, while if placed in series with a resistance 495.05 ohms it gives a full-scale deflection for 10 volts. Deduce the resistance of the galvanometer and the current required to produce a full-scale deflection when it is used alone. (L.)

2. A rectangular coil of wire of n turns and area A is suspended at the mid-point of one side by a fibre of torsional constant c so that its plane is parallel to a horizontal uniform field of magnetic induction (flux density) B . Derive an expression for the deflection of the coil when a steady current I flows through it.

Explain, with reasons, how a modern moving coil galvanometer of the suspended type has been developed from this simple arrangement.

Distinguish between the accuracy and sensitivity of such a galvanometer and explain the factors on which the sensitivity depends. (N.)

3. Describe a moving-coil type of galvanometer and deduce a relation between its deflection and the steady current passing through it.

A galvanometer, with a scale divided into 150 equal divisions, has a current sensitivity of 10 divisions per milliampere and a voltage sensitivity of 2 divisions per millivolt. How can the instrument be adapted to serve (a) as an ammeter reading to 6 A, (b) as a voltmeter in which each division represents 1 volt? (L.)

4. Describe an experiment to show that a force is exerted on a conductor carrying a current when it is placed in a magnetic field. Give a diagram showing the directions of the current, the field, and the force.

A rectangular coil of 50 turns hangs vertically in a uniform magnetic field of 10^{-2} T (Wb m^{-2}), so that the plane of the coil is parallel to the field. The mean height of the coil is 5 cm and its mean width 2 cm. Calculate the strength of the current that must pass through the coil in order to deflect it 30° if the torsional constant of the suspension is 10^{-9} newton metre per degree. Give a labelled diagram of a moving-coil galvanometer. (L.)

5. Explain the terms *magnetic flux*, *magnetic flux density (magnetic induction)*, *magnetic flux linkage*.

Write down an expression for the force experienced by a long straight wire of length L carrying a current I when it is situated at right angles to a uniform field of flux-density B . State the units in which each of the quantities in the expression is measured, and show with the aid of a diagram the directions of the current, the magnetic field and the force.

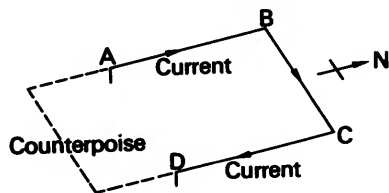


FIG. 35.23.

Fig. 35.23 represents a rigid rectangular wire frame with three members $AB = 30$ cm, $BC = 60$ cm, and $CD = 30$ cm. It is pivoted at the contacts A and D

and suitably counterpoised. When BC is at right angles to the magnetic meridian and a current of 5 A passes along ABCD as shown, the frame is found to be in equilibrium when its plane is horizontal. Find the magnitude and direction of the force on each of the members AB, BC, and CD, and also the moment of the deflecting couple which results. (Take the flux density of the earth's magnetic field to be 6×10^{-5} T (Wb m^{-2}) and the angle of dip to be 70° .)

State, and explain briefly without attempting any numerical calculation, what would be obtained (a) if the current through the frame were reversed, (b) if the whole system were turned through 90° about a vertical axis, the counterpoising couple in each case remaining the same as before. (O.)

6. Define *magnetic moment*.

A magnetized uniform rod of length 0.15 m is pivoted about a horizontal axis passing through its centre of mass. The axis of rotation is perpendicular to a uniform magnetic field of 64 A m^{-1} inclined at 60° to the horizontal. When a small mass of 0.2 g is fixed to one end of the rod, the rod sets horizontally. Draw a diagram to show the couples acting on the rod in this position. Calculate the magnetic moment of the rod. (N.)

7. Draw a clear labelled diagram showing the essential features of a moving-coil ballistic galvanometer. What differences in structure would be necessary and for what reasons, to make the galvanometer dead beat?

Describe, with a circuit diagram, how a ballistic galvanometer may be used to compare the capacitances of two capacitors and state the precautions necessary to obtain a reliable result. (L.)

8. Describe the construction of a sensitive moving-coil galvanometer. How could the instrument be adapted for use a millivoltmeter?

A standard cell of e.m.f. 1.018 volts and internal resistance 1000 ohms is joined to two resistances in series of values 149 000 and 2 ohms respectively. The ends of the 2-ohm resistance are also connected to the terminals of a galvanometer of resistance 8 ohms, when a scale deflection of 100 mm is recorded. What is the sensitivity of the instrument expressed in microamperes per scale division? (L.)

9. With the help of a labelled diagram or diagrams describe the construction and explain the action of a pivoted moving-coil galvanometer. Indicate on diagrams how the direction of deflection is related to the direction of current flow and the polarity of the magnet.

If such an instrument has a resistance of 10 ohms and gives a full-scale deflection when a current of 20 milliamps is passing, how would it be converted into a voltmeter with 3- and 150-volt ranges? What reading would this voltmeter give when connected to a battery of e.m.f. 120 volts and internal resistance 300 ohms? (L.)

chapter thirty-six

Electromagnetic Induction

Faraday's Discovery

AFTER Ampere and others had investigated the magnetic effect of a current Faraday attempted to find its converse: he tried to produce a current by means of a magnetic field. He began work on the problem in 1825 but did not succeed until 1831.

The apparatus with which he worked is represented in Fig. 36.1; it consists of two coils of insulated wire, A, B, wound on a wooden core.

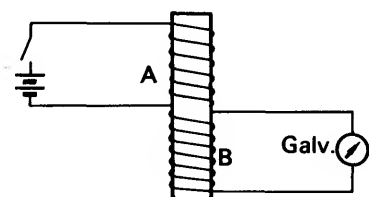


FIG. 36.1. Faraday's experiment on induction.

One coil was connected to a galvanometer, and the other to a battery. No current flowed through the galvanometer, as in all Faraday's previous attempts. But when he disconnected the battery Faraday happened to notice that the galvanometer needle gave a kick. And when he connected the battery back again, he noticed a kick in the opposite direction. However often he disconnected and reconnected the battery, he got the same results. The 'kicks' could hardly be all accidental—they must indicate momentary currents. Faraday had been looking for a steady current, but the effect he sought turned out to be a transient one—that was why it took him six years to find it.

Conditions for Generation of Induced Current

The results of Faraday's experiments showed that a current flowed in coil B of Fig. 36.1 only while the magnetic field due to coil A was changing—the field building up as the current in A was switched on, decaying as the current in A was switched off. And the current which flowed in B while the field was decaying was in the opposite direction to the current which flowed while the field was building up. Faraday called the current in B an induced current. He found that it could be made much greater by winding the two coils on an iron core, instead of a wooden one.

Once he had realized that an induced current was produced only by a change in the magnetic field inducing it, Faraday was able to find induced currents wherever he had previously sought them. In place of the coil A he used a magnet, and showed that as long as the coil and the magnet were at rest, there was no induced current (Fig. 36.2 (i)). But when he moved either the coil or the magnet an induced current flowed as long as the motion continued (Fig. 36.2 (ii)). If the current flowed one way when the north pole of the magnet was approaching

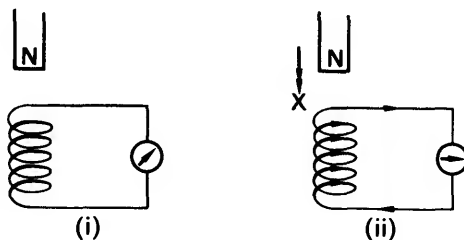


FIG. 36.2. Induction of current by moving magnet.

the end X of the coil, it flowed the other way when the north pole was retreating from X, or the south pole approached X.

Since a flow of current implies the presence of an e.m.f., Faraday's experiments showed that an e.m.f. could be induced in a coil by moving it relatively to a magnetic field. In discussing induction it is more fundamental to deal with the e.m.f. than the current, because the current depends on both the e.m.f. and the resistance.

Direction of E.M.F.; Lenz's Law

Before considering the magnitude of an induced e.m.f., let us investigate its direction. To do so we must first see which way the galvanometer deflects when a current passes through it in a known direction: we can find this out with a battery and a megohm resistor (Fig. 36.3 (i)). We then take a coil whose direction of winding we know, and connect this to the galvanometer. In turn we plunge each pole of a magnet into and out of the coil; and we get the results shown in Fig. 36.3 (ii), (iii), (iv). These results were generalized most elegantly into a rule by Lenz in 1835. He said that *the induced current flows always in such a direction as to oppose the change which is giving rise to it*. If the reader will sketch with a pencil on Fig. 36.3 the magnetic fields of the induced currents, then he will see what Lenz meant: when the magnet is approaching the coil, the coil repels it; when the magnet is retreating from the coil, the coil attracts it.

Lenz's law is a beautiful example of the conservation of energy: the induced current sets up a force on the magnet, which the mover of the

magnet must overcome: the work done in overcoming this force provides the electrical energy of the current. (This energy is dissipated as heat in the coil.) If the induced current flowed in the opposite direction to

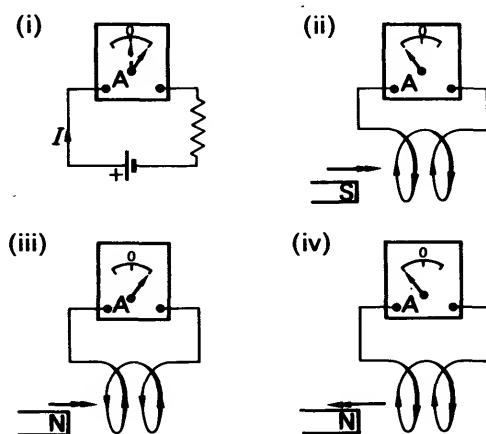


FIG. 36.3. Direction of induced currents.

that which it actually takes, then it would aid—it would speed up—the motion of the magnet. It would enhance its own cause, and grow indefinitely; at the same time, it would continuously increase the kinetic energy of the magnet. Thus both mechanical and electrical energy would be produced, without any agent having to do work. The system would be a perpetual motion machine.

The direction of the induced e.m.f., E , is specified by that of the current, as in Fig. 36.4. If we wished to reword Lenz's law, substituting e.m.f. for current, we would have to speak of the e.m.f.s *tending* to oppose the change... etc., because there can be no opposing force unless the circuit is closed and a current can flow.

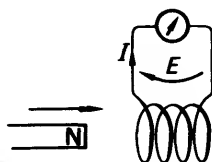


FIG. 36.4. Direction of induced e.m.f.

Magnitude of E.M.F.

Accurate experiments on induction are difficult to contrive with simple apparatus; but rough-and-ready experiments will show on what factors the magnitude depends. We require coils of the same diameter but different numbers of turns, coils of the same number of turns but different diameters, and two similar magnets, which we can use singly or together. If we use a high-resistance galvanometer, the current will not vary much with the resistance of the coil in which the e.m.f. is induced, and we can take the deflection as a measure of the e.m.f. There is no need to plunge the magnet into and out of the coil: we can

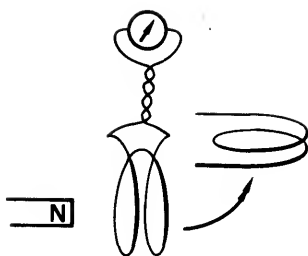


FIG. 36.5. E.m.f. induced by turning coil.

get just as great a deflection by simply turning the coil through a right angle, so that its plane changes from parallel to perpendicular to the magnet, or vice versa (Fig. 36.5). We find that the induced e.m.f. increases with:

- (i) the speed with which we turn the coil;
- (ii) the area of the coil;
- (iii) the strength of the magnetic field (two magnets give a greater e.m.f. than one);

(iv) the number of turns in the coil.

To generalize these results and to build up useful formulae, we use the idea of *magnetic flux*, or field lines passing through a coil. Fig. 36.6 shows a coil, of area A , whose normal makes an angle θ with a uniform magnetic field of induction B . The component of the field at right angles to the plane of the coil is $B \cos \theta$, and we say that the magnetic flux Φ through the coil is

$$\Phi = AB \cos \theta \quad . \quad . \quad . \quad (1)$$

(We get the same result if we multiply the field-strength B by the area projected at right angles to the field, $A \cos \theta$.) If either the strength of the field is changed, or the coil is turned so as to change the angle θ , then the flux through the coil changes.

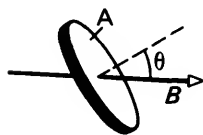


FIG. 36.6.
Magnetic flux.

Results (i) to (iii) above, therefore, show that the e.m.f. induced in a coil increases with the *rate of change of the magnetic flux* through it. More accurate experiments show that the induced e.m.f. is actually proportional to the rate of change of flux through the coil; this result is sometimes called *Faraday's*, or *Neumann's*, law.

The unit of magnetic flux Φ is called the *weber* (Wb). Hence the unit of B is the *weber per metre*² (Wb m⁻²) or *tesla* (T).

Flux Linkage

If a coil has more than one turn, then the flux through the whole coil is the sum of the fluxes through the individual turns. We call this the *flux linkage* through the whole coil. If the magnetic field is uniform, the flux through one turn is given, from (1), by $AB \cos \theta$. If the coil has N turns, the total flux linkage Φ is given by

$$\Phi = NAB \cos \theta \quad . \quad . \quad . \quad (2)$$

From Faraday's or Neumann's law, the e.m.f. induced in a coil is proportional to the rate of change of the flux linkage, Φ . Hence

$$E \propto \frac{d\Phi}{dt},$$

or

$$E = -k \frac{d\Phi}{dt}, \quad (3)$$

where k is a positive constant. The minus sign expresses Lenz's law. It means that the induced e.m.f. is in such a direction that, if the circuit is closed, the induced current *opposes* the change of flux. Note that an induced e.m.f. exists across the terminals of a coil when the flux linkage changes, even though the coil is on 'open circuit'. A current, of course, does not flow in the latter case.

On p. 901, it is shown that $E = -kd\Phi/dt$ is consistent with the expression $F = BIl$ for the force on a conductor only if $k = 1$. We may therefore say that

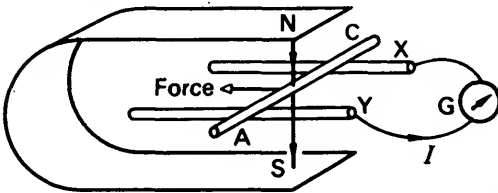
$$E = -\frac{d\Phi}{dt}, \quad (4)$$

where Φ is the flux linkage in webers, t is in seconds, and E is in volts.

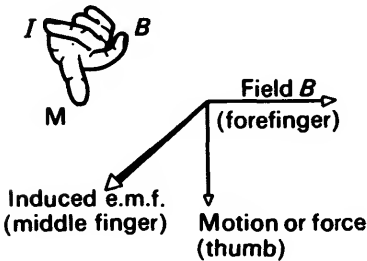
From (4), it follows that one weber is the flux linking a circuit if the induced e.m.f. is one volt when the flux is reduced uniformly to zero in one second.

E.M.F. Induced in Moving Rod

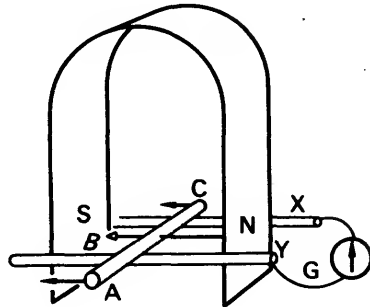
Generators at power stations produce high induced voltages by rotating long *straight conductors*. Fig. 36.7 (i) shows a simple apparatus



(i) Demonstration



(iii) Direction



(ii) No e.m.f.

FIG. 36.7. E.m.f. induced in moving rod.

for demonstrating that an e.m.f. may be induced in a straight rod or wire, when it is moved across a magnetic field. The apparatus consists of a rod AC resting on rails XY, and lying between the poles NS of a permanent magnet. The rails are connected to a galvanometer G.

If we move the rod to the left, so that it cuts across the field B of the magnet, a current I flows as shown. If we move the rod to the right, the current reverses. We notice that the current flows only while the rod is moving, and we conclude that the motion of the rod AC induces an e.m.f. in it.

By turning the magnet into a vertical position (Fig. 36.7 (ii)) we can show that no e.m.f. is induced in the rod when it moves parallel to the field B . We conclude that an e.m.f. is induced in the rod only when it *cuts across* the field. And, whatever the direction of the field, no e.m.f. is induced when we slide the rod parallel to its own length. The induced e.m.f. is greatest when we move the rod at right angles, both to its own length and to the magnetic field. These results may be summarized in *Fleming's right-hand rule*:

If we extend the thumb and first two fingers of the right hand, so that they are all at right angles to one another, then the directions of field, motion, and induced e.m.f. are related as in Fig. 36.7 (iii).

To show $E.M.F. \propto \text{Rate of Change}$

The variation of the magnitude of the e.m.f. in a rod with the speed of 'cutting' magnetic flux can be demonstrated with the apparatus in Fig. 36.8 (i).

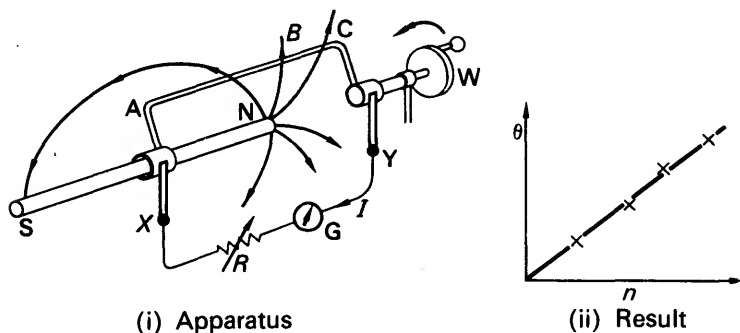


FIG. 36.8. Induced e.m.f.

Here AC is a copper rod, which can be rotated by a wheel W around one pole N of a long magnet. Brushing contacts at X and Y connect the rod to a galvanometer G and a series resistance R. When we turn the wheel, the rod AC cuts across the field B of the magnet, and an e.m.f. is induced in it. If we turn the wheel steadily, the galvanometer gives a steady deflection, showing that a steady current is flowing round the circuit.

To find how the current and e.m.f. depends on the speed of the rod, we keep the circuit resistance constant, and vary the rate at which we

turn the wheel. We time the revolutions with a stop-watch, and find that the deflection θ is proportional to the number of revolutions per second, n (Fig. 36.8 (ii)). It follows that the induced e.m.f. is proportional to the speed of the rod.

Calculation of E.M.F. in Rod

Consider the circuit shown in Fig. 36.9. PQ is a straight wire touching the two connected parallel wires QR, PS and free to move over them. All the conductors are situated in a uniform magnetic field of induction B , perpendicular to the plane of PQRS.

Suppose the rod PQ is pulled with a uniform velocity v by an external force F . There will then be a change of flux linkage in the area PQRS and so an e.m.f. will be induced in the circuit. This produces a current I which flows round the circuit. A force will now act on the wire PQ

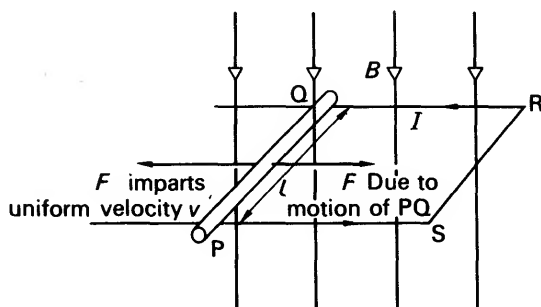


FIG. 36.9 Calculation of induced e.m.f.

due to the current flowing and to the presence of the magnetic field (p. 878). By Lenz's law, the direction of this force will oppose the movement of PQ. If the current flowing is I , and the length of PQ is l , the force on PQ is BIl . This is equal to the external force F , since PQ is not accelerating.

Because energy is conserved, the rate of working by the external force is equal to the rate at which energy is supplied to the electrical circuit. Now in one second, PQ moves a distance v . Hence

work done per second = force \times distance moved per second.

$$= BIlv$$

If the induced e.m.f. is E , the electrical energy used in one second, or power, = El .

$$\therefore EI = BIlv$$

$$\therefore E = Blv. \quad (6)$$

This result has been derived without using the relation $E = -d\Phi/dt$. To see if the same result as (6) can be obtained, consider the flux changes. In one second the area of PQRS changes by vl . Hence the

change in flux linkage per second, $d\Phi/dt = B \times \text{area change per second} = Blv$. Hence, numerically,

$$\therefore E = Blv.$$

This means that the relation $E = -d\Phi/dt$ may be used to find the induced e.m.f. in a straight wire.

Induced E.M.F. and Force on Moving Electrons

We have seen that an electron moving across a magnetic field experiences a mechanical force (p. 881). This explains neatly the e.m.f. induced in a wire: when we move the wire across the field, we move each free electron in it, likewise across the field. As Fig. 36.10 shows, the force on the electrons, F , is at right angles to the plane containing the velocity v of the wire, and the magnetic field B . Thus it tends to drive the electrons along the wire. The direction in which it does so agrees with the direction of the conventional e.m.f., which is the direction of the force on a positive charge.

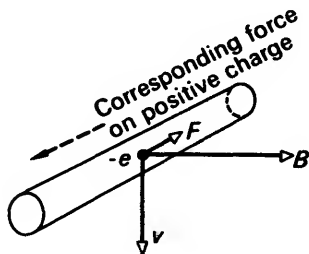


FIG. 36.10. Force on a moving electron.

When a wire AC is swept, as shown in Fig. 36.11, across a magnetic field B , the force on the electrons in it acts from A to C. Therefore, if the wire is not connected to a closed circuit, electrons will pile up at C: the end C will gain a negative

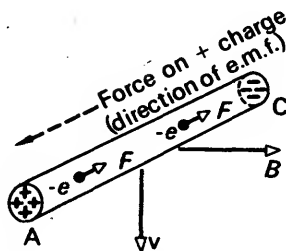


FIG. 36.11. Induced e.m.f. arising from force on moving electrons.

charge, and A will be left with a positive charge. The end A will therefore be at a higher potential than C.

If we now clear our minds of electrons, we see that the conventional e.m.f. acts from C to A. If positive electricity were free to move, it would accumulate at A;

in other words, the tendency of the e.m.f., acting from C to A, is to give A a higher potential than C. The *potential difference* between A and C tends to drive positive electricity the other way. Equilibrium is reached when the potential difference V_{AC} is equal to the e.m.f. acting from C to A.

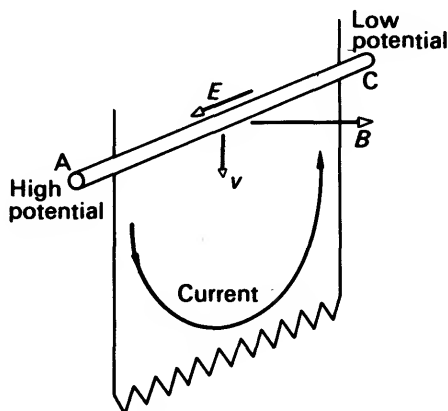


FIG. 36.12. E.m.f. and potential difference.

When the wire is connected to a closed circuit, current flows from A to C round the external circuit (Fig. 36.12). Within the source of current—the wire AC—the e.m.f. drives the current (of positive charge) from C to A: from low potential to high. *This is the essential function of an e.m.f.*; an e.m.f. is an agency which can drive an electric current *against* a potential difference. When the e.m.f. arises in a wire moving across a magnetic field, this agency is the force on the electrons moving with the wire.

The e.m.f. induced in a wire can easily be calculated from the force on a moving electron. If the wire moves with a velocity v at right angles to a field B , then so do the electrons in it. Each of them therefore experiences a force

$$F = Bev,$$

(equation (1), p. 881), where e is the electronic charge. The work which this force does in carrying the electron along the length l of the wire is Fl . But it is also, by definition, equal to the product of the e.m.f. E , and the charge e . Therefore

$$Ee = Fl = Bev l,$$

whence

$$E = Blv.$$

APPLICATIONS OF INDUCTION

The Induction Coil

The induction coil is a device for getting a high voltage from a low one. It was at one time used for X-ray tubes (p. 1067), and is nowadays used in car radios. It consists of a core of iron wires, around which is wrapped a coil of about a hundred turns of thick insulated wire, called the primary (Fig. 36.13 (i)). Around the primary is wound the secondary coil, which has many thousands of turns of fine insulated wire. The

primary is connected to a battery of accumulators, via a make-and-break M , which works in the same way as the contact-breaker of an electric bell: it switches the current on and off many times a second, thus varying the magnetic flux.

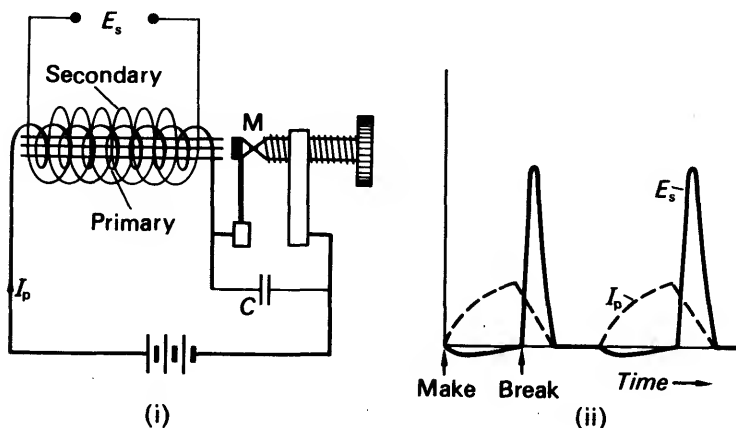


FIG. 36.13. Induction coil.

When the primary current I_p is switched on, the rise of its magnetic field induces an e.m.f. E_s in the secondary. A similar e.m.f., but in the opposite sense, is induced in the secondary when the primary current is switched off, by the collapse of the magnetic field. The secondary e.m.f.s are determined by the number of turns in the secondary coil, and by the rate of change of the magnetic flux through the iron core. Because of the great number of secondary turns, the secondary e.m.f.s may be high and of the order of thousands of volts (Fig. 36.13 (ii)).

In practice, an induction coil such as we have described—consisting simply of primary, secondary, and contact-breaker—would not give high secondary e.m.f.s. For, at the make of the primary current, the current would rise slowly, because of the self-inductance (see p. 924) of the primary winding. The rate of change of flux linked with the secondary would therefore be small, and the secondary e.m.f. low. And at the break of the primary current a spark would pass between the contacts of the make-and-break. The spark would allow primary current to continue to flow, and the primary current would fall slowly. At the instant of break, before the spark began, the primary current would be falling rapidly and the secondary e.m.f. would be high; but the e.m.f. would remain high for only a very short time: as soon as the spark passed the secondary e.m.f. would fall to a value about as low as at make.

Nothing can be done about the low secondary e.m.f. at make. But the secondary e.m.f. at break can be made high, by preventing sparking at the contact-breaker. To prevent sparking, a capacitor, C in Fig. 36.13, is connected across the contacts.

As we shall see on p. 926, the capacitor actually *slows down* the fall of the primary current at the instant of break; but in doing so it prevents the induced e.m.f. in the primary from rising high enough to start a spark. And the rate at which the primary current falls, in charging the capacitor, is greater than the rate at which it would fall if a spark were passing. Thus, with a capacitor, the secondary e.m.f. is less at the instant of break than without one, but it is greater throughout the rest of the fall of the primary current. Consequently the average secondary voltage at break is higher with a capacitor than without; in practice it is much higher. To get the greatest possible secondary voltage, the capacitance of the capacitor is chosen so that it just suppresses sparking at the contacts. The secondary voltage is then a series of almost unidirectional pulses, as shown in Fig. 36.13 (ii).

The iron core of an induction coil is made from a bundle of wires, to minimize eddy-currents (p. 913). If eddy-currents were to flow they would, by Lenz's law, set up a flux opposing the change of primary current. Thus they would reduce the secondary e.m.f.

The Dynamo and Generator

Faraday's discovery of electromagnetic induction was the beginning of electrical engineering. Nearly all the electric current used today is generated by induction, in machines which contain coils moving continuously in a magnetic field.

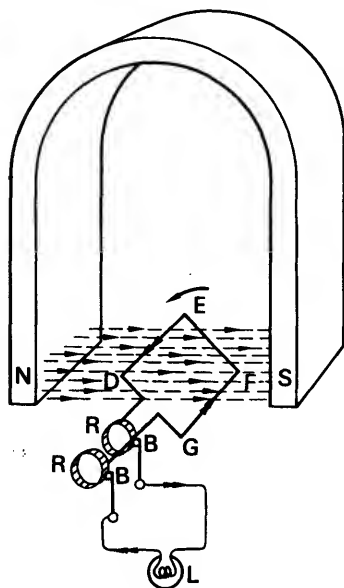


FIG. 36.14. A simple dynamo.

Fig. 36.14 illustrates the principle of such a machine, which is called a dynamo, or generator. A coil DEFG, shown for simplicity as having only one turn, rotates on a shaft, which is not shown, between the poles NS of a horseshoe magnet. The ends of the coil are connected to flat brass rings R, which are supported on the shaft by discs of insulating material, also not shown. Contact with the rings is made by small blocks of carbon B, supported on springs, and shown connected to a lamp L. As the coil rotates, the flux linking it changes, and a current is induced in it which flows, via the carbon blocks, through the lamp. The magnitude (which we study shortly) and the direction of the current are not constant. Thus when the coil is in the position shown, the limb ED is moving downwards through the lines of force, and GF

is moving upwards. Half a revolution later, ED and GF will have interchanged their positions, and ED will be moving upwards. Consequently, applying Fleming's right-hand rule, the current round the coil must reverse as ED changes from downward to upward motion. The

actual direction of the current at the instant shown on the diagram is indicated by the double arrows, from Fleming's rule. By applying this rule, it can be seen that the current reverses every time the plane of the coil passes the vertical position.

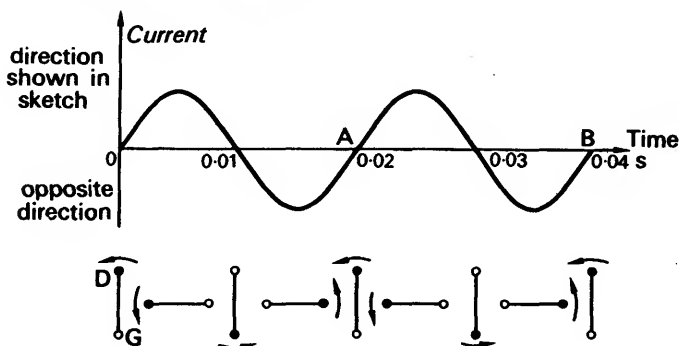


FIG. 36.15. Current generated by dynamo of Fig. 36.14, plotted against time and coil position.

We shall see shortly that the magnitude of the e.m.f. and current varies with time as shown in Fig. 36.15; this diagram also shows the corresponding position of DG. This type of current is called an *alternating current* (A.C.). A complete alternation, such as from A to B in the figure, is called a 'cycle'; and the number of cycles which the current goes through in one second is called its 'frequency'. The frequency of the current represented in the figure is that of most domestic supplies in Britain—50 Hz (cycles per second).

E.M.F. in Dynamo

We can now calculate the e.m.f. in the rotating coil. If the coil has an area A , and its normal makes an angle θ with the magnetic field B , as in Fig. 36.16, then the flux through the coil

$$= AB \cos \theta \text{ (see p. 898).}$$

The flux linkages with the coil, if it has N turns, are

$$\Phi = NAB \cos \theta.$$

If the coil turns with a steady angular velocity ω or $d\theta/dt$, then the e.m.f. induced in volts in the coil is

$$\begin{aligned} E &= -\frac{d\Phi}{dt} \\ &= -NAB \frac{d}{dt}(\cos \theta) \\ &= NAB \sin \theta \frac{d\theta}{dt} \end{aligned} \quad (1)$$

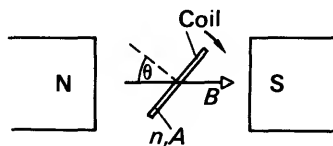


FIG. 36.16. Coil inclined to magnetic field.

In terms of the number of revolutions per second, f , which the coil makes, we have

$$\frac{d\theta}{dt} = 2\pi f$$

and

$$\theta = 2\pi ft,$$

$$\therefore E = 2\pi f NAB \sin 2\pi ft. \quad (2)$$

Thus the e.m.f. varies sinusoidally with time, like the pressure in a sound-wave, the frequency being f cycles per second.

The maximum (peak) value or amplitude of E occurs when $\sin 2\pi ft$ reaches the value 1. If the maximum value is denoted by E_0 , it follows that

$$E_0 = 2\pi f NAB,$$

and

$$E = E_0 \sin 2\pi ft. \quad (3)$$

The e.m.f. E sends an alternating current, of a similar sine equation, through a resistor connected across the coil.

Alternators

Generators of alternating current are often called *alternators*. In all but the smallest, the magnetic field of an alternator is provided by an electromagnet called a field-magnet or *field*, as shown in Fig. 36.17; it has a core of cast steel, and is fed with direct current from a separate d.c.

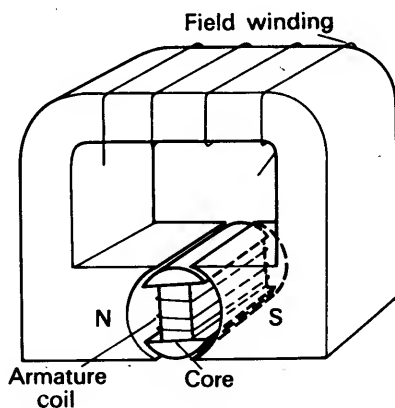


FIG. 36.17. Field magnet and armature.

generator. The rotating coil, called the armature, is wound on an iron core, which is shaped so that it can turn within the pole-pieces of the field-magnet. With the field-magnet, the armature core forms a system which is almost wholly iron, and can be strongly magnetized by a small current through the field winding. The field in which the armature turns is much stronger than if the coil had no iron core, and the e.m.f. is

proportionately greater. In the small alternators used for bicycle lighting the armature is stationary, and the field is provided by permanent magnets, which rotate around it. In this way rubbing contacts, for leading the current into and out of the armature, are avoided.

When no current is being drawn from a generator, the horse-power required to turn its armature is merely that needed to overcome friction, since no electrical energy is produced. But when a current is drawn, the horse-power required increases, to provide the electrical power. The current, flowing through the armature winding, causes the magnetic field to set up a couple which opposes the rotation of the armature, and so demands the extra horse-power. The reader should check the truth of this statement by marking the direction of the e.m.f., current, and force on the limbs of the coil in Fig. 36.17.

The Transformer

A transformer is a device for stepping up—or down—an alternating voltage. It has primary and secondary windings, as in an induction coil, but no make-and-break (Fig. 36.18). It has an iron core, which is made from E-shaped laminations, interleaved so that the magnetic flux does not pass through air at all; in this way the greatest flux is obtained

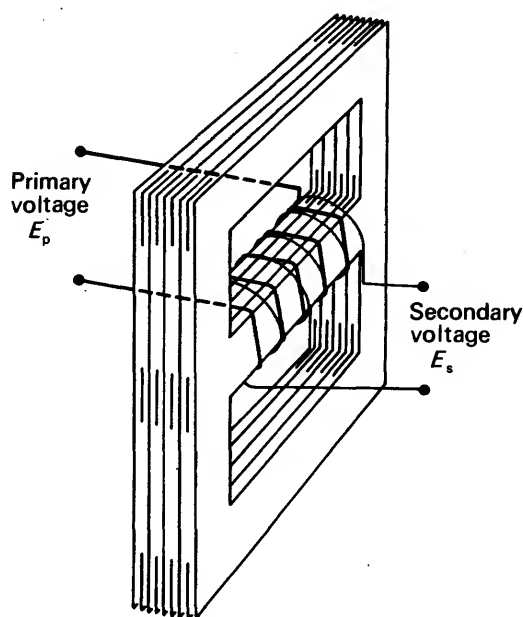


FIG. 36.18. Transformer.

with a given current. When an alternating e.m.f. E_p is impressed on the primary winding, it sends an alternating current through it, which sets up an alternating flux in the core of magnitude BA , where B is

the induction and A is the cross-sectional area. This induces an alternating e.m.f. in the secondary E_s . If N_p , N_s are the number of turns in the primary and secondary coils, their linkages with the flux Φ are:

$$\Phi_p = N_p AB \quad \Phi_s = N_s AB$$

The magnitude of the e.m.f. induced in the secondary is, from the formula on p. 899:

$$E_s = \frac{d\Phi_s}{dt} = N_s A \frac{dB}{dt}$$

The changing flux also induces a back-e.m.f. in the primary, whose magnitude is

$$E_p = \frac{d\Phi_p}{dt} = N_p A \frac{dB}{dt}$$

Because the primary winding has inevitably some resistance, the current flowing through it sets up a voltage drop across the resistance. But in practice this is negligible compared with the back-e.m.f. due to the changing flux. Consequently we may say that the voltage applied to the primary, from the source of current, is used simply in overcoming the back-e.m.f. E_p . Therefore it is equal in magnitude to E_p . (This is analogous to saying, in mechanics, that action and reaction are equal and opposite.) Consequently we have

$$\frac{\text{e.m.f. induced in secondary}}{\text{voltage applied to primary}} = \frac{E_s}{E_p} = \frac{N_s}{N_p} \quad (1)$$

Thus the transformer steps voltage up or down according to its 'turns-ratio':

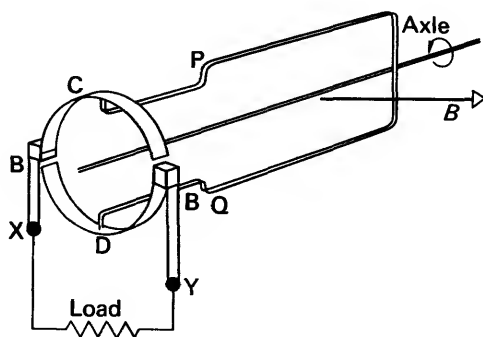
$$\frac{\text{secondary voltage}}{\text{primary voltage}} = \frac{\text{secondary turns}}{\text{primary turns}}$$

When a load is connected to the secondary winding, a current flows in it. This current flows in such a direction as to reduce the flux in the core. At the instant that the load is connected, therefore, the back-e.m.f. in the primary falls. The primary current then increases. The increase in primary current increases the flux through the core, and continues until the flux is restored to its original value. The back-e.m.f. in the primary is then again equal to the applied voltage, and equilibrium is restored. But now a greater primary current is flowing than before the secondary was loaded. Thus the power drawn from the secondary is drawn, in turn, from the supply to which the primary is connected.

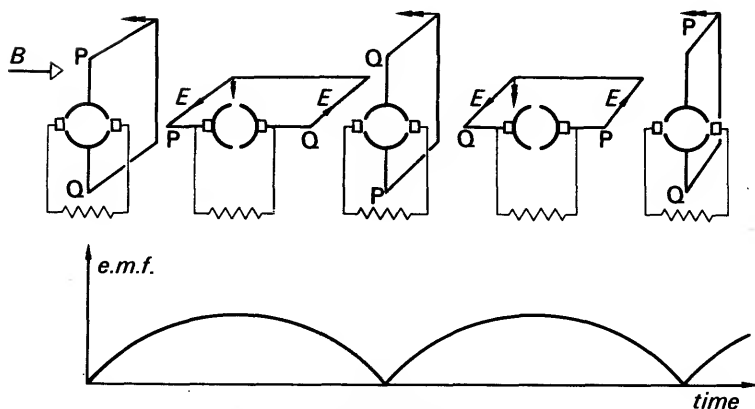
Transformers are used to step up the voltage generated at a power station, from 11000 to 132 000 volts for high-tension transmission (p. 793). After transmission they are used to step it down again to a value safer for distribution (240 volts in houses). Inside a house a transformer may be used to step the voltage down from 240 to 4, for ringing bells. Transformers with several secondaries are used in, for example, radio-receivers, where several different voltages are required.

D.C. Generators

Fig. 36.19 (i) is a diagram of a direct-current generator or dynamo. Its essential difference from an alternator is that the armature winding is connected to a *commutator* instead of slip-rings. The commutator consists of two half-rings of copper C, D, insulated from one another, and turning with the coil. Brushes BB, with carbon tips, press against the commutator and are connected to the external circuit. The commutator



(i) Principle



(ii) Output

FIG. 36.19. D.C. generator.

is oriented so that it reverses the connexions from the coil to the circuit at the instant when the e.m.f. reverses in the coil. Fig. 36.19 (ii) shows several positions of the coil and commutator, and the e.m.f. observed at the terminals XY. This e.m.f. pulsates in magnitude, but it acts always in the same sense round the circuit connected to XY. It is a pulsating direct e.m.f. The average value in this case can be shown to be $2/\pi$ of the maximum e.m.f. E_0 , given in equation (3), p. 907.

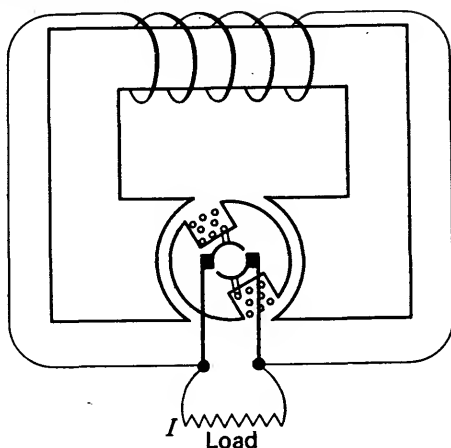


FIG. 36.20. D.C. generator with energized field.

In practice, as in an alternator, the armature coil is wound with insulated wire on a soft iron core, and the field-magnet is energized by a current (Fig. 36.20). This current is provided by the dynamo itself. The steel of the field-magnet has always a small residual magnetism, so that as soon as the armature is turned an e.m.f. is induced in it. This then sends a current through the field winding, which increases the field and the e.m.f.; the e.m.f. rapidly builds up to its working value.

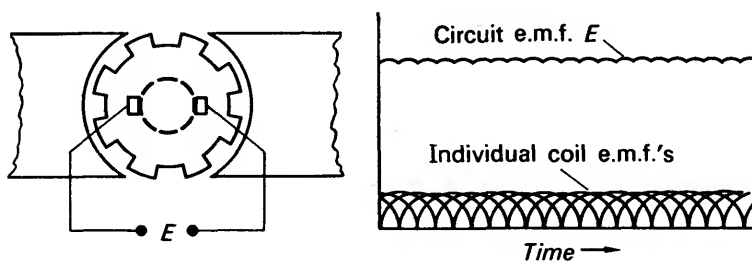


FIG. 36.21. D.C. generator with segment commutator.

Most consumers of direct current wish it to be steady, not pulsating as in Fig. 36.19. A reasonably steady e.m.f. is given by an armature with many coils, inclined to one another, and a commutator with a correspondingly large number of segments. The coils are connected to the commutator in such a way that their e.m.f.s add round the external circuit (Fig. 36.21).

Homopolar Generator

Another type of generator, which gives a very steady e.m.f., is illustrated in Fig. 36.22. It consists of a copper disc which rotates between

the poles of a magnet; connexions are made to its axle and circumference. If we assume (as is not true) that the magnetic field B is uniform over the radius XY , then we can calculate the induced e.m.f. E . In one revolution the radius XY sweeps out an area $(\pi(r_1^2 - r_2^2))$, where r_1 and r_2 are the radii of the wheel and the axle. If T is the time for one revolution, then the rate at which XY sweeps out area is $\pi(r_1^2 - r_2^2)/T$.

The rate at which it sweeps out flux is therefore

$$\frac{\pi(r_1^2 - r_2^2)}{T} B = \pi(r_1^2 - r_2^2) B f,$$

where f denotes the revolutions of the wheel per second. Thus

$$\begin{aligned} E &= \pi(r_1^2 - r_2^2) B f \\ &= \pi(r_1^2 - r_2^2) B f \end{aligned}$$

Generators of this kind are called *homopolar* because the e.m.f. induced in the moving conductor is always in the same sense. They are sometimes used for electroplating, where only a small voltage is required, but they are not useful for most purposes, because they give too small an e.m.f. The e.m.f. of a commutator dynamo can be made large by having many turns in the coil; but the e.m.f. of a homopolar dynamo is limited to that induced in one radius of the disc.

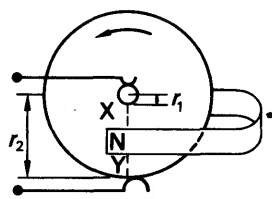


FIG. 36.22.
E.m.f. induced in a disc.

Applications of Alternating and Direct Currents

Direct currents are less easy to generate than alternating currents, and alternating e.m.f.s are more convenient to step up and to step down, and to distribute over a wide area. The national grid system, which supplies electricity to the whole country, is therefore fed with alternating current. Alternating current is just as suitable for heating as is direct current, because the heating effect of a current is independent of its direction. It is also equally suitable for lighting, because filament lamps depend on the heating effect, and gas-discharge lamps—neon, sodium, mercury—run as well on alternating current as on direct. Small motors, of the size used in vacuum-cleaners and common machine-tools, run satisfactorily on alternating current, but large ones, as a general rule, do not. Direct current is therefore used on most electric railway and tramway systems. These systems either have their own generating stations, or convert alternating current from the grid into direct current. One way of conveying alternating current into direct is to use a valve rectifier, whose principle we shall describe later.

For electro-chemical processes alternating current is useless. The chemical effect of a current reverses with its direction, and if, therefore, we tried to deposit a metal by alternating current, we would merely cause a small amount of the metal to be alternately deposited and

dissolved. For electroplating, and for battery charging, alternating current must be rectified.

Eddy-currents

The core of the armature of a dynamo is built up from thin sheets of soft iron insulated from one another by an even thinner film of oxide, as shown in Fig. 36.23 (i). These are called laminations, and the armature is said to be laminated. If the armature were solid, then, since iron is a conductor, currents would be induced in it by its motion across the magnetic field (Fig. 36.23 (ii)). These currents would absorb power by opposing the rotation of the armature, and they would dissipate that power as heat, which would damage the insulation of the winding; but

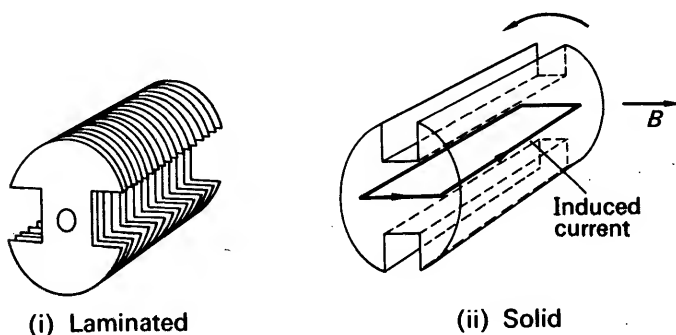


FIG. 36.23. Armature cores.

when the armature is laminated, these currents cannot flow, because the induced e.m.f. acts at right angles to the laminations, and therefore to the insulation between them. The magnetization of the core, however, is not affected, because it acts along the laminations. Thus the eddy-currents are suppressed, while the desired e.m.f.—in the armature coil—is not.

Eddy-currents, by Lenz's law, always tend to oppose the motion of a solid conductor in a magnetic field. The opposition can be shown in many ways. One of the most impressive is to make a chopper with a thick copper blade, and to try to slash it between the poles of a strong electromagnet; then to hold it delicately and allow it to drop between them. The resistance to the motion in the former case can be felt.

Sometimes eddy-currents can be made use of—for example, in damping a galvanometer. When a current is passed through the coil of a galvanometer, it applies a couple to the coil which sets it swinging. If the swings are opposed only by the viscosity of the air, they decay very slowly and are said to be naturally damped (Fig. 36.24). The pointer or light-spot takes a long time to come to its final steady deflection θ . To bring the spot or pointer more rapidly to rest, the damping must be increased. One way of increasing the damping is to wind the coil on a metal former. Then, as the coil swings, the field of

the permanent magnet induces eddy-currents in it; and these, by Lenz's law, oppose its motion. They therefore slow down the turning of the coil towards its eventual position, but they also suppress its swings about that position; in the end the coil comes to rest sooner than if it

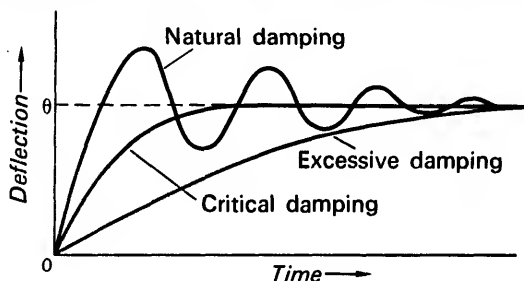


FIG. 36.24. Damping of galvanometer.

were not damped. Galvanometer coils which are wound on insulating formers can be damped by short-circuiting a few of their turns, or by connecting an external shunt across the whole coil. With a shunt the eddy-currents circulate round the coil and shunt, independently of the current to be measured. The smaller the shunt, the greater the eddy-currents and the damping; if the coil is overdamped, as shown in Fig. 36.24, it may take almost as long to come to rest as when it is undamped. The damping which is just sufficient to prevent overshoot is called 'critical' damping.

Electric Motors

If a simple direct-current dynamo, of the kind described on p. 910, is connected to a battery it will run as a motor (Fig. 36.25). Current flows round the armature coil, and the magnetic field exerts a couple on this, as in a moving-coil galvanometer. The commutator reverses the current just as the limbs of the coil are changing from upward to

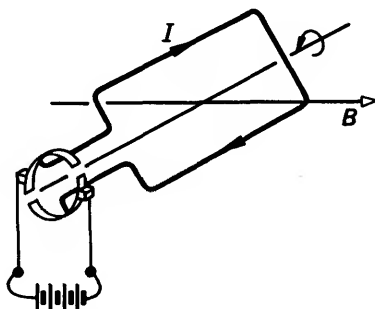


FIG. 36.25. Principle of D.C. motor.

downward movement and vice versa. Thus the couple on the armature is always in the same sense, and the shaft turns continuously. (The reader should verify these statements with the help of Fig. 36.25.)

The armature of a motor is laminated, in the same way and for the same reason, as the armature of a dynamo.

Back-e.m.f. in Motor

When the armature of a motor rotates, an e.m.f. is induced in its windings; by Lenz's law this e.m.f. opposes the current which is making the coil turn. It is therefore called a back-e.m.f. If its magnitude is E , and V is the potential difference applied to the armature by the supply, then the armature current is

$$I_a = \frac{V - E}{R_a}. \quad (1)$$

Here R_a is the resistance of the armature, which is generally small—of the order of 1 ohm.

The back-e.m.f. E is proportional to the strength of the magnetic field, and the speed of rotation of the armature. When the motor is first switched on, the back-e.m.f. is zero: it rises as the motor speeds up. In a large motor the starting current would be ruinously great; to limit it, a variable resistance is inserted in series with the armature, which is gradually reduced to zero as the motor gains speed.

When a motor is running, the back-e.m.f. in its armature E is not much less than the supply voltage V . For example, a motor running off the mains ($V = 230$ volts) might develop a back-e.m.f. $E = 220$ volts. If the armature had a resistance of 1 ohm, the armature current would then be 10 amp (equation (1)). When the motor was switched on, the armature current would be 230 amp if no starting resistor were used.

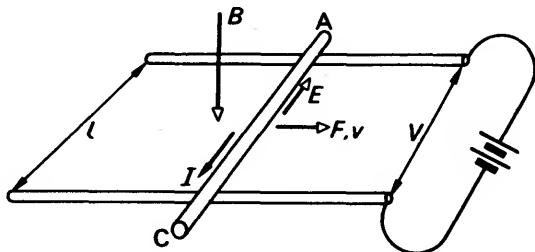


FIG. 36.26. Back-e.m.f. and mechanical power.

Back-e.m.f. and Power

The back-e.m.f. in the armature of a motor represents the mechanical power which it develops. To see that this is so, we use an argument similar to that which we used in finding an expression for the e.m.f. induced in a conductor. We consider a rod AC, able to slide along rails, in a plane at right angles to a magnetic field B (Fig. 36.26). But we now suppose that a current I is maintained in the rod by a battery, which sets up a potential difference V between the rails. The magnetic field then exerts a force F on the rod, given by

$$F = BIl.$$

The force F makes the rod move; if its velocity is v , the mechanical power developed by the force F is

$$P_m = Fv = BIlv \quad (1)$$

As the rod moves, a back-e.m.f. is induced in it, whose magnitude E is given by the expression for the e.m.f. in a moving rod (p. 901):

$$E = Blv. \quad (2)$$

Equations (1) and (2) together give

$$P_m = EI. \quad (3)$$

Thus the mechanical power developed is equal to the product of the back-e.m.f. and the current.

Before returning to consider motors, we may complete the analysis of the action represented in Fig. 36.26. If R is the resistance of the rails and rod, the heat developed in them is I^2R . The power supplied by the battery is IV , and the battery is the only source of power in the whole system. Therefore

$$IV = I^2R + P_m; \quad (4)$$

the power supplied by the battery goes partly into heat, and partly into useful mechanical power. Also, by equation (3),

$$IV = I^2R + EI, \quad (5)$$

whence

$$V = IR + E$$

or

$$I = \frac{V - E}{R}.$$

This is equation (1), p. 915, which we previously obtained simply from Ohm's law.

Let us apply this theory to the example which we were considering. We had:

supply voltage, $V = 230$ volts;

back-e.m.f., $E = 220$ volts;

armature resistance, $R_a = 1$ ohm;

armature current, $I_a = 10$ amp.

The power dissipated as heat in the armature is $I_a^2 R_a = 100 \times 1 = 100$ watts. The power supplied to the armature is $I_a V = 10 \times 230 = 2300$ watts, and the mechanical power is $I_a E = 10 \times 220 = 2200$ watts. Of the power supplied to the armature, the fraction which appears as mechanical power is $2200/2300 = 96$ per cent. This is not, however, the efficiency of the motor as a whole, because current is taken by the winding on the field magnet.

The Field Winding

The field winding of a motor may be connected in series or in parallel with the armature. If it is connected in series, it carries the armature

current, which is large (Fig. 36.27). The field winding therefore has few turns of thick wire, to keep down its resistance, and so the power wasted in it as heat. The few turns are enough to magnetize the iron, because the current is large. If the field coil is connected in parallel with the armature, as in Fig. 36.28, the motor is said to be 'shunt-wound'. The field winding has many turns of fine wire to keep down

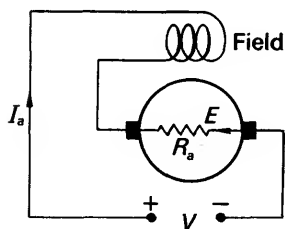


FIG. 36.27. Series-wound motor.

the current which it consumes. In the above example, if the motor is shunt-wound and the field current is 0.5 A, then the power dissipated as heat in the field is $0.5 \times 230 = 115$ watts. The power consumption of the motor is therefore $2300 + 115 = 2415$ watts, and its efficiency is

$$\frac{\text{mechanical power developed}}{\text{electrical power consumed}} = \frac{2220}{2415} = 92 \text{ per cent.}$$

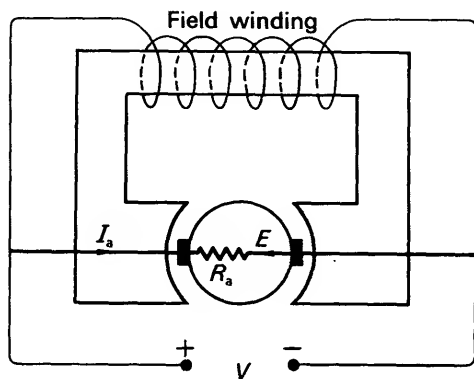


FIG. 36.28. Current and voltages in shunt-wound motor.

The working efficiency of the motor will be a little less than this, because some of the mechanical power will be used in overcoming friction in the bearings.

Shunt Field

Shunt-wound motors are used for driving machine-tools, and in other jobs where a steady speed is required. A shunt motor keeps a nearly steady speed for the following reason. If the load is increased, the speed falls a little; the back-e.m.f. then falls in proportion to the speed, and the current rises, enabling the motor to develop more power to overcome the increased load. In the example, p. 916, if the speed falls by 1 part in 220, the back-e.m.f. falls from 220 to 219 V. The current then rises from $\frac{230-220}{1} = 10$ A to $\frac{230-219}{1} = 11$ A. And the mechanical power increases from $220 \times 10 = 2200$ watts to $219 \times 11 = 2409$ watts (≈ 2400). Thus an increase in load of $\frac{2400-2200}{2200} = 9$ per cent causes a fall in speed of 1 part in 220—less than $\frac{1}{2}$ per cent.

Series Field

Series motors are used where great torque is required in starting—for example, in cranes.

They develop a great starting torque because the armature current flows through the field coil. At the start the armature back-e.m.f. is small, and the current is great—as great as the starting resistance will allow. The field-magnet is therefore very strongly magnetized. The torque on the armature is proportional to the field and to the armature current; since both are great at the start, the torque is very great.

A series motor does not keep such a steady speed as a shunt motor. Just as in a shunt motor, when the load increases the speed falls; and the fall in speed decreases the back-e.m.f., and allows more current to flow. But, as we will see in a moment, the back-e.m.f. in a series motor, does not fall with the speed as sharply as it does in a shunt motor. To meet a given increase in load, the armature current must increase by a definite amount. And therefore the back-e.m.f. must fall by a definite amount. But it falls less with the speed than it does in a shunt motor. Consequently, to meet a given increase in load, the speed of a series motor must fall more than that of a shunt motor.

We now show that the back-e.m.f. in a series motor falls less with the speed than in a shunt one. The argument is best given in steps:

- (i) when the speed falls, the back-e.m.f. falls;
- (ii) the current through both armature and field winding increases;
- (iii) the field becomes stronger;
- (iv) the increase in the field tends to *increase* the back-e.m.f., i.e. to offset its initial fall;
- (v) thus the very fall of the back-e.m.f., by permitting a greater current and strengthening the field, tends to offset itself;
- (vi) therefore the back-e.m.f. falls slowly with the speed—more slowly than in a shunt motor, where the field is constant;
- (vii) as we have already seen, this means that the speed must fall further, to meet a given increase in load.

CHARGE AND FLUX LINKAGE

Flux and Charge relation

We have already seen that an electromotive force is induced in a circuit when the magnetic flux linked with it changes. If the circuit is closed, a current flows, and electric charge is carried round the circuit. As we shall now show, there is a simple relationship between the charge and the change of flux.

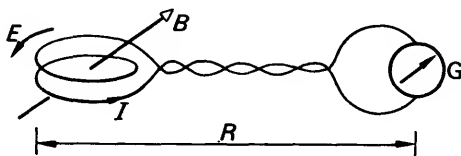


FIG. 36.29. Coil with changing flux.

Consider a closed circuit of total resistance R ohm, which has a total flux linkage Φ with a magnetic field (Fig. 36.29). If the flux

linkages start to change,

$$\text{induced e.m.f., } E = -\frac{d\Phi}{dt}.$$

$$\therefore \text{current, } I = \frac{E}{R} = -\frac{1}{R} \frac{d\Phi}{dt} \quad (1)$$

In general, the flux linkage will not change at a steady rate, and the current will not be constant. But, throughout its change, charge is being carried round the circuit. If a time t seconds is taken to reach a new constant value, the charge carried round the circuit in that time is

$$Q = \int_0^t I dt.$$

From (1),

$$\begin{aligned} \therefore Q &= -\frac{1}{R} \int_0^t \frac{d\Phi}{dt} \cdot dt \\ &= -\frac{1}{R} \int_{\Phi_0}^{\Phi_t} d\Phi, \end{aligned}$$

where Φ_0 is the number of linkages at $t = 0$, and Φ_t is the number of linkages at time t . Thus

$$Q = -\frac{\Phi_t - \Phi_0}{R} = \frac{\Phi_0 - \Phi_t}{R}.$$

The quantity $\Phi_0 - \Phi_t$ is positive if the linkages Φ have decreased, and negative if they have increased. But as a rule we are interested only in the magnitude of the charge, and we may write

$$Q = \frac{\text{change of flux linkage}}{R} \quad (2)$$

Equation (2) shows that the charge circulated is proportional to the change of flux-linkages, and independent of the time.

Ballistic Galvanometer

It can be seen from the last section that the charge which flows round a given circuit is directly proportional to the change of flux linkage. If the charge flowing is measured by a *ballistic galvanometer* G , as shown in Fig. 36.29, then we have a measure of the change in flux linkage, Φ .

Ballistics is the study of the motion of a body, such as a projectile, which is set off by a blow, and then allowed to move freely. By freely, we mean without friction. A ballistic galvanometer is one used to measure an electrical blow, or impulse: for example, the charge Q which circulates when a capacitor is discharged through it. A galvanometer which is intended to be used ballistically has a heavier coil than one which is not; and it has as little damping as possible—an insulating former, no short-circuited turns, no shunt. The mass of its coil makes it swing slowly; in the example above, for instance, the capacitor has

discharged, and the charge has finished circulating, while the galvanometer coil is just beginning to turn. The galvanometer coil continues to turn, however; and as it does so it twists the suspension. The coil stops turning when its kinetic energy, which it gained from the forces set up by the current, has been converted into potential energy of the suspending fibre. The coil then swings back, as the suspension untwists itself, and it continues to swing back and forth for some time. Eventually it comes to rest, but only because of the damping due to the viscosity of the air, and to the internal friction of the fibre. Theory shows that, if the damping is negligible, *the first deflection of the galvanometer is proportional to the quantity of electricity, Q , that passed through its coil, as it began to move.* This first deflection, θ , is often called the 'throw' of the galvanometer; we have, then,

$$Q = k\theta, \quad (1)$$

where k is a constant of the galvanometer.

Equation (1) is true only if all the energy given to the coil is spent in twisting the suspension. If an appreciable amount of energy is used to overcome damping—i.e. dissipated as heat by eddy currents—then the galvanometer is not ballistic, and θ is not proportional to Q .

To calibrate the ballistic galvanometer, a capacitor of known capacitance, e.g. $2 \mu\text{F}$, is charged by a battery of known e.m.f., e.g. 50 volt, and then discharged through the instrument. See p. 768. Suppose the deflection is 200 divisions. The charge $Q = CV = 100$ microcoulomb, and thus the galvanometer sensitivity is 2 divisions per microcoulomb.

Measurement of Induction

Fig. 36.30 illustrates the principle of measuring the induction B in the field between the poles of a powerful magnet. A small coil, called a *search coil*, with a known area and number of turns, is connected to a ballistic galvanometer G . It is positioned at right angles to the field to be measured, as shown, so that the flux enters the coil face normally.

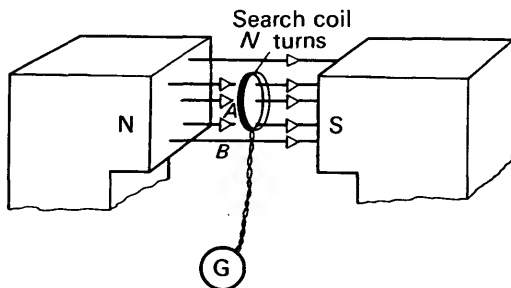


FIG. 36.20. Induction by ballistic galvanometer.

The coil is then pulled completely out of the field by moving it smartly downwards, for example, and the throw θ produced in the galvanometer is observed. The charge q which passes round the circuit is proportional to θ , from above.

Suppose B is the field-strength in Wb m^{-2} or tesla (T), A is the area of the coil in m^2 and N is the number of turns. Then

$$\text{change of flux-linkages} = NAB$$

$$\therefore \text{quantity, } Q, \text{ through galvanometer} = \frac{NAB}{R},$$

where R is the *total* resistance of the galvanometer and search coil. But

$$Q = c\theta,$$

where c is the quantity per unit deflection of the ballistic galvanometer.

$$\begin{aligned} \therefore \frac{NAB}{R} &= c\theta \\ \therefore B &= \frac{Rc\theta}{NA}. \end{aligned} \quad (1)$$

The constant c is found by discharging a capacitor through the galvanometer (see p. 768). If C is the capacitance in farads, V the p.d. in volts of the battery originally charging it, and α the deflection of the galvanometer, then $c = CV/\alpha$ coulomb per unit deflection.

The Earth Inductor

As another example of the use of a search coil and ballistic galvanometer, we describe a method that has been used for measuring the angle of dip of the earth's magnetic field (p. 944). The earth's field is so nearly uniform that the search coil may be large, usually about 30 cm square; but the field is also so weak that even a large coil must have many turns—of the order of 100. The coil, which is called an earth inductor, is pivoted in a wooden frame, and this is fitted with stops so that the coil can be turned rapidly through 180° (Fig. 36.31). To find

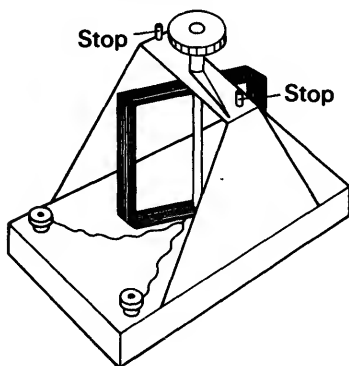


FIG. 36.31. Earth inductor.

the angle of dip, we connect the coil to a ballistic galvanometer, and set it with its plane horizontal, as shown at (i) in Fig. 36.32. The flux linking the coil at N turns is then

$$\Phi = NAB_v,$$

where A is the area of the coil, and B_v is the vertical component of the earth's field. If we were to turn the coil through 90° , the flux would fall to zero; and if we were to turn it through a further 90° , the flux linkage would become NAB_v once more, but it would thread the coil in the opposite direction. Therefore we turn the coil through 180° , and change the flux linkage by $2NAB_v$; at the same time we observe the throw, θ , of the galvanometer.

If R is the total resistance of galvanometer and search coil, the circulated charge is, by equation (2) on p. 919,

$$Q = \frac{\Phi}{R} = \frac{2NAB_v}{R}.$$

But

$$Q = c\theta,$$

where c is the constant of the galvanometer. Therefore

$$\frac{2NAB_v}{R} = c\theta. \quad (1)$$

We now set the frame of the earth inductor so that the axis of the coil is vertical, and so that, when the coil is held by one of the stops, its plane lies East–West. See Fig. 36.32 (ii). The flux threading the coil is now NAB_H , where B_H is the horizontal component of the earth's field. Therefore, when we turn the coil through 180° , the throw θ' of the galvanometer is given by

$$\frac{2NAB_H}{R} = c\theta' \quad (2)$$

Now the angle of dip, δ , is given by

$$\tan \delta = \frac{B_v}{B_H}.$$

Therefore, from equations (1) and (2),

$$\tan \delta = \frac{\theta}{\theta'}.$$

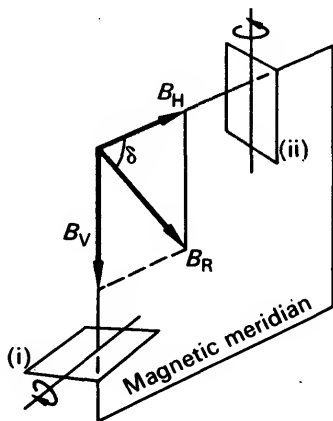


FIG. 36.32. Measurement of dip.

Self-induction

The phenomenon which we call self-induction was discovered by the American, Joseph Henry, in 1832. He was led to it by a theoretical argument, starting from the phenomena of induced e.m.f., which he had discovered at about the same time as Faraday.

When a current flows through a coil, it sets up a magnetic field. And that field threads the coil which produces it. Fig. 36.33 (i). If the current through the coil is changed—by means of a variable resistance, for example—the flux linked with the turn of the coil changes. An e.m.f. is therefore induced in the coil. By Lenz's law the direction of the induced e.m.f. will be such as to oppose the change of current; the e.m.f. will be

against the current if it is increasing, with it if it is decreasing (Fig. 36.33 (ii)).

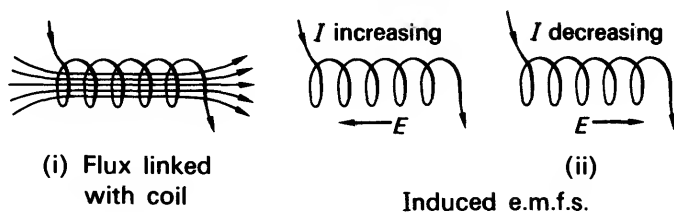


FIG. 36.33. Self-induction.

Back-E.M.F.

When an e.m.f. is induced in a circuit by a change in the current through that circuit, the process of induction is called self-induction. The e.m.f. induced is called a back-e.m.f. Self-induction opposes the growth of current in a coil, and so makes it gradual. This effect can be demonstrated by connecting an iron-cored coil of many turns in series with an ammeter and a few accumulators (Fig. 36.34 (i)). (The ammeter

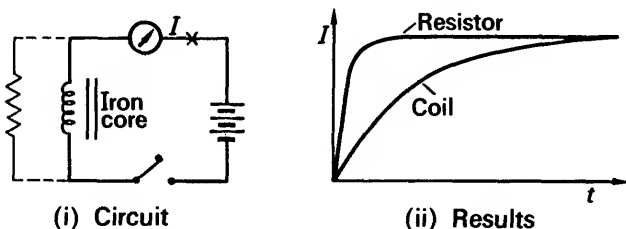


FIG. 36.34. Self-induction experiment.

should be of the 'short-period' type and critically damped.) When the current is switched on, the pointer of the ammeter moves slowly over to its final position. If the coil is now replaced by a rheostat of the same resistance, the pointer moves much more swiftly to the same reading (Fig. 36.34 (ii)).

Just as self-induction opposes the rise of an electric current when it is switched on, so also it opposes the decay of the current when it is switched off. When the circuit is broken, the current starts to fall very rapidly, and a correspondingly great e.m.f. is induced, which tends to maintain the current. This e.m.f. is often great enough to break down the insulation of the air between the switch contacts, and produce a spark. To do so, the e.m.f. must be about 350 volts or more, because air will not break down—not over any gaps, narrow or wide—when the voltage is less than that value. The e.m.f. at break may be much greater than the e.m.f. of the supply which maintained the current: a spark can easily be obtained, for example, by breaking a circuit consisting of an iron-cored coil and an accumulator.

Non-inductive Coils

In bridge circuits, such as are used for resistance measurements, self-induction is a nuisance. When the galvanometer key of a bridge is closed, the currents in the arms of the bridge are redistributed, unless the bridge happens to be balanced. While the currents are being redistributed they are changing, and self-induction delays the reaching of a new equilibrium. Thus the galvanometer deflection at the instant of closing the key, does not correspond to the steady state which the bridge will eventually reach. It may therefore be misleading. To minimize their self-inductance, the coils of bridges and resistance boxes are wound so as to set up extremely small magnetic fields: as shown in

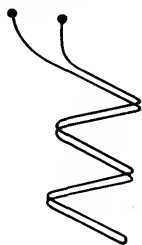


FIG. 36.35.
Non-inductive
winding.

Fig. 36.35, the wire is doubled-back on itself before being coiled up. Every part of the coil is then traversed by the same current travelling in opposite directions, and its magnetic field is negligible. Such a coil is said to be non-inductive.

When describing the use of a bridge, we said that the battery key should be pressed before the galvanometer key. Doing so gives time for the currents in the arms of the bridge to become steady before the galvanometer key is pressed. It therefore minimizes any possible effects of self-induction.

Self-inductance

To discuss the effects of self-induction we must define the property of a coil which gives rise to them. This property is called the self-inductance of the coil, and is defined as follows:

$$\text{self-inductance} = \frac{\text{back-e.m.f. induced in coil by a changing current}}{\text{rate of change of current through coil}}$$

Self-inductance is denoted by the symbol L ; we may therefore write its definition as

$$L = \frac{E_{\text{back}}}{dI/dt}$$

or

$$E_{\text{back}} = L \frac{dI}{dt} \quad (1)$$

Equation (1) is the simplest form in which to remember the definition.

The unit of self-inductance is the henry (H). It is defined by making each term in equation (1) equal to unity; thus a coil has a self-inductance of 1 henry if the back-e.m.f. in it is 1 volt, when the current through it is changing at the rate of 1 ampere per second. Equation (1) then becomes:

$$E_{\text{back}} \text{ (volts)} = L \text{ (henrys)} \times \frac{dI}{dt} \text{ (ampere/second).}$$

The iron-cored coils used for smoothing the rectified supply current to a radio receiver (p. 1011) are usually very large and have an inductance of about 30 henrys.

***L* for Coil**

Since the induced e.m.f. $E = d\Phi/dt = L = dI/dt$, numerically, it follows by integration from a limit of zero that

$$\Phi = LI.$$

Thus $L = \Phi/I$. Hence the self-inductance may be defined as the *flux linkage per unit current*. When Φ is in webers and I in amperes, then L is in henrys. Thus if a current of 2A produces a flux linkage of 4 Wb in a coil, the inductance $L = 4 \text{ Wb}/2\text{A} = 2\text{H}$.

We shall see later that when a long coil of N turns and length l carries a current I , (i) the magnetizing field $H = NI/l$ and (ii) this produces a flux-density or induction B inside the coil given by $B = \mu H$, where μ is the permeability of the material inside the coil (p. 941). Hence

$$\text{flux linkage } \Phi = NAB = NA\mu H = \frac{\mu N^2 AI}{l}$$

$$\therefore L = \frac{\Phi}{I} = \frac{\mu N^2 A}{l} \quad (1)$$

This formula may be used to find the approximate value of the inductance of a coil. L is in henrys when A is in metre², l in metre and μ is in henry metre⁻¹.

Energy Stored; E.M.F. at Break

The spark which passes when the current in a coil is interrupted liberates energy in the form of heat and light. This energy has been stored in the magnetic field of the coil, just as the energy of a charged capacitor is stored in the electrostatic field between its plates (p. 779). When the current in the coil is first switched on, the back-e.m.f. opposes the rise of current; the current flows against the back-e.m.f. and therefore does work against it (p. 795). When the current becomes steady, there is no back-e.m.f. and no more work done against it. The total work done in bringing the current to its final value is stored in the magnetic field of the coil. It is liberated when the current collapses; for then the induced e.m.f. tends to maintain the current, and to do external work of some kind.

To calculate the energy stored in a coil, we suppose that the current through it is rising at a rate dI/dt ampere per second. Then, if L is its self-inductance in henrys, the back-e.m.f. across it is given by

$$E = L \frac{dI}{dt} \text{ volt.}$$

If the value of the current, at the instant concerned, is I amperes, then the rate at which work is being done against the back-e.m.f. is

$$P = EI = LI \frac{dI}{dt} \text{ watt.}$$

The total work done in bringing the current from zero to a steady value I_0 is therefore

$$W = \int P dt = \int_0^{I_0} LI \frac{dI}{dt} dt = \int_0^{I_0} LI dI \\ = \frac{1}{2} LI_0^2 \text{ joule.}$$

This is the energy stored in the coil.

To calculate the e.m.f. induced at break is, in general, a complicated business. But we can easily do it for one important practical circuit. To prevent sparking

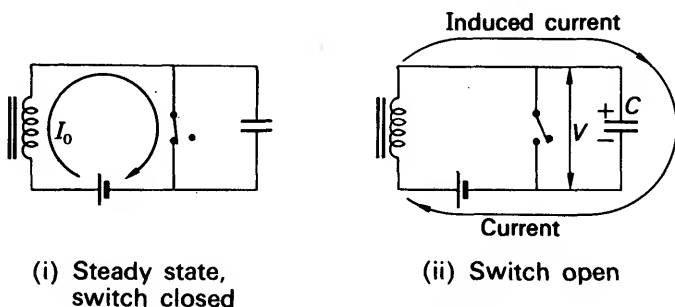


FIG. 36.36. Prevention of sparking by capacitor.

at the contacts of the switch in an inductive circuit, a capacitor is often connected across them (Fig. 36.36 (i)). When the circuit is broken, the collapsing flux through the coil tends to maintain the current; but now the current can continue to flow for a brief time; it can flow by charging the capacitor (Fig. 36.36 (ii)). Consequently the current does not decay as rapidly as it would without the capacitor, and the back-e.m.f. never rises as high. If the capacitance of the capacitor is great enough, the potential difference across it (and therefore across the switch) never rises high enough to cause a spark.

To find the value to which the potential difference does rise, we assume that all the energy originally stored in the magnetic field of the coil is now stored in the electrostatic field of the capacitor.

If C is the capacitance of the capacitor in farad, and V_0 the final value of potential difference across it in volt, then the energy stored in it is $\frac{1}{2} CV_0^2$ joule (p. 779). Equating this to the original value of the energy stored in the coil, we have

$$\frac{1}{2} CV_0^2 = \frac{1}{2} LI_0^2.$$

Let us suppose that a current of 1 ampere is to be broken, without sparking, in a circuit of self-inductance 1 henry. To prevent sparking, the potential difference across the capacitor must not rise above 350 volt. The least capacitance that must be connected across the switch is therefore given by

$$\frac{1}{2} C \times 350^2 = \frac{1}{2} \times 1 \times 1^2.$$

Hence
$$C = \frac{1}{350^2} = 8 \times 10^{-6} \text{ farad} = 8 \mu\text{F}.$$

A paper capacitor of capacitance $8 \mu\text{F}$, and able to withstand 350 volts, would therefore be required.

Mutual Induction

We have already seen that an e.m.f. may be induced in one circuit by a changing current in another (Fig. 36.1, p. 895). The phenomenon is often called mutual induction, and the pair of circuits which show it are said to have mutual inductance. The mutual inductance, M , between two circuits is defined by the equation:

$$\left. \begin{array}{l} \text{e.m.f. induced in B, by} \\ \text{changing current in A} \end{array} \right\} = M \times \left\{ \begin{array}{l} \text{rate of change of} \\ \text{current in A.} \end{array} \right.$$

See Fig. 36.37. In symbols,

$$E_B = M \frac{dI_A}{dt}.$$

Mutual inductance is truly mutual; it is the same from B to A as A to B. Its unit is the same as that of self-inductance, the henry.

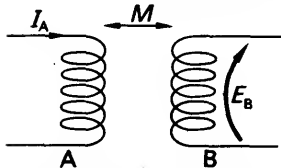


FIG. 36.37. Mutual induction.

EXERCISES 36

1. State Lenz's law of electromagnetic induction and describe, with explanation, an experiment which illustrates its truth.

Describe the structure of a transformer suitable for supplying 12 volts from 240-volt mains and explain its action. Indicate the energy losses which occur in the transformer and explain how they are reduced to a minimum.

When the primary of a transformer is connected to the a.c. mains the current in it (a) is very small if the secondary circuit is open, but (b) increases when the secondary circuit is closed. Explain these facts. (L.)

2. Describe, with the aid of a large labelled diagram, the structure of a simple form of a.c. generator.

Explain (a) how it could be modified to produce direct current;

(b) the features that enable it to produce a high e.m.f. (compared with a cell);

(c) the features that minimize the heat wasted. (L.)

3. Define *electromotive force* and state the *laws of electromagnetic induction*. Using the definition and the laws, derive an expression for the e.m.f. induced in a conductor moving in a magnetic field.

When a wheel with metal spokes 120 cm long is rotated in a magnetic field of flux density $0.5 \times 10^{-4} \text{ Wb m}^{-2}$ normal to the plane of the wheel, an e.m.f. of 10^{-2} volt is induced between the rim and the axle. Find the rate of rotation of the wheel. (L.)

4. Describe the differences in structure and action between a non-ballistic and a ballistic moving-coil galvanometer.

A corrected deflection of 24 scale divisions of a ballistic galvanometer is obtained *either* by charging a capacitor of $3 \mu\text{F}$ capacitance to a potential difference of 2 volt and discharging it through the galvanometer, or by connecting the ballistic galvanometer in series with a flat circular coil of 80 turns each of diameter 1 cm, the combined resistance of coil and galvanometer being 2000 ohms, and quickly thrusting the coil into a strong magnetic field so that the plane of the coil is perpendicular to the direction of the field. State the sensitivity of the galvanometer and calculate the strength of the magnetic field. (The strength of the earth's magnetic field may be neglected.) (L.)

5. Explain what is meant by *self inductance* and define the practical unit in which it is measured.

Describe and explain an experiment which demonstrates the phenomenon of self-induction. (N.)

6. State the laws relating to (a) the direction, (b) the magnitude of an electromagnetically induced electromotive force, and describe very briefly an experiment illustrating each.

Deduce the relation between the quantity of electricity flowing through a circuit and the flux change producing it.

A flat coil of 150 turns, each of area 300 cm^2 and of total resistance 50 ohms, is connected to a circuit whose resistance is 40 ohms. Starting with its plane horizontal, the coil is rotated quickly through a half-turn about a diametral axis pointing along the magnetic meridian. If the quantity of electricity which then flows round the circuit is 4 microcoulombs, find the intensity of the vertical component of the earth's magnetic field. (N.)

7. What are eddy currents?

Describe and explain an experiment in which eddy currents are produced.

Describe one useful application of eddy currents. (N.)

8. Define *self inductance* and *mutual inductance*.

Explain the differences in structure and action between a ballistic and an aperiodic galvanometer.

A ballistic galvanometer of resistance 15 ohms and sensitivity 5 divisions per microcoulomb is connected in series with a resistance of 100 ohms and a secondary coil of 500 turns and of resistance 50 ohms. This coil is wound round the middle of a long solenoid of radius 3 cm having 10 turns cm^{-1} and carrying a current of 0.6 A. Assuming no damping, calculate the deflection produced in the galvanometer when the current in the solenoid is switched off. (L.)

9. Describe experiments (one in each case) involving the use of a moving-coil ballistic galvanometer to (a) compare two capacitances of approximately the same magnitude, (b) compare the magnetic induction (flux density) between the poles of one electromagnet with that between the poles of another. In each case justify the method used to calculate the result.

Explain two special features of a galvanometer suitable for use in these experiments. (N.)

10. State Lenz's law and describe how you would demonstrate it using a solenoid with two separate superimposed windings with clearly visible turns, a cell with marked polarity, and a centre-zero galvanometer. Illustrate your answer with diagrams.

A metal aircraft with a wing span of 40 m flies with a ground speed of 1000 km h^{-1} in a direction due east at constant altitude in a region of the northern hemisphere where the horizontal component of the earth's magnetic field is $1.6 \times 10^{-5} \text{ T (Wb m}^{-2}\text{)}$ and the angle of dip is 71.6° . Find the potential difference in volts that exists between the wing tips and state, with reasons, which tip is at the higher potential. (N.)

11. State the laws of electromagnetic induction and describe briefly experiments to show their validity.

A coil A passes a current of 1.25 A when a steady potential difference of 5 V is maintained across it, and an r.m.s. current of 1 A when it has across it a sinusoidal potential difference of 5 V r.m.s. at a frequency of 60 Hz (cycles per second). Explain why the current is less in the second case, and calculate the resistance and the inductance of the coil.

The same coil A , which has 100 turns, has a second coil B with 500 turns wound on it so that all the magnetic flux produced by A is linked by B . Find the r.m.s. value of the e.m.f. that appears across the open-circuit ends of B when a sinusoidal alternating current of 1 A r.m.s. at a frequency of 50 Hz is passed through A . Why is the ratio of this e.m.f. to the r.m.s. potential difference across A not the same as the ratio of the number of turns in B and A , i.e. 5:1?

Explain why the insertion of an iron core into the coils would decrease the current in A and increase the e.m.f. across B , if the alternating potential difference across A were kept unchanged; the effects of hysteresis and eddy currents in the iron may be neglected. ($O. \& C.$)

12. A choke of large self inductance and small resistance, a battery and a switch are connected in series. Sketch and explain a graph illustrating how the current varies with time after the switch is closed. If the self inductance and resistance of the coil are 10 henrys and 5 ohms respectively and the battery has an e.m.f. of 20 volts and negligible resistance, what are the greatest values after the switch is closed of (a) the current, (b) the rate of change of current? ($N.$)

13. State the laws of electromagnetic induction, and describe an experiment by which one of them can be verified.

A piece of wire 8 cm long, of resistance 0.020 ohm and mass 22 mg, is bent to form a closed square $ABCD$. It is mounted so as to turn without friction about a horizontal axis through AB ; a uniform horizontal magnetic field of flux-density 0.50 Wb m^{-2} is applied at right angles to this axis. The side CD is raised until the plane of the square is horizontal and then released. Calculate approximately the time taken for the plane of the square to become vertical. You can assume that, during the falling, the couple due to gravity is equal and opposite to that due to electromagnetic forces. ($C.$)

14. State the laws of electromagnetic induction. Hence derive an expression for the time variation of the electromotive force induced in a single turn of wire rotating about an axis in its plane, the axis being perpendicular to a uniform magnetic field. Explain the action of a simple alternating current generator.

What modification of this generator is required to produce a direct current? Indicate by a sketch how the e.m.f. across the output terminals of a single coil would vary with time in the case of (a) an a.c. and (b) a d.c. generator. How may a more uniform output e.m.f. be obtained in the latter case?

A rectangular coil of wire having 100 turns, of dimensions $30 \text{ cm} \times 30 \text{ cm}$, is rotated at a constant speed of 600 r.p.m. in a magnetic field of 0.1 Wb m^{-2} , the axis of rotation being in the plane of the coil and perpendicular to the field. Calculate the induced e.m.f. ($L.$)

15. Describe with the aid of diagrams (a) a transformer and (b) an induction coil. Explain the action of each.

Draw diagrams to show, in a general way, how the voltage output from each of these appliances varies with the time. ($N.$)

16. What are eddy-currents? Give two examples of the practical use of such currents.

A metal disc of diameter 20 cm rotates at a constant speed of 600 r.p.m. about an axis through its centre and perpendicular to its plane in a uniform magnetic field of $5 \times 10^{-3} \text{ Wb m}^{-2}$ established parallel to the axis of rotation. Calculate the e.m.f. in volts between the centre and rim of the disc. Show clearly on a diagram the direction of rotation of the disc and the direction of the magnetic field and of the e.m.f. induced. ($L.$)

17. How would you show that a change in the number of lines of magnetic force, however produced, threading through a circuit produces an induced e.m.f.?

A magnet is suspended by a thin wire so that its axis is horizontal and its centre is above the centre of a circular copper disc, mounted horizontally. Explain what happens to the magnet when the disc is rotated.

What would be the effect of replacing the disc by one of identical dimensions but made of a substance of high resistivity? (L.)

18. State Lenz's law and describe fully a method by which you could verify this law experimentally.

A horizontal metal disc of radius 10 cm is rotated about a central vertical axis at a region where the value of the earth's magnetic flux density is $5.3 \times 10^{-5} \text{ T}$ (Wb m^{-2}) and the angle of dip is 70° . A sensitive galvanometer of resistance 150 ohms is connected between the centre of the disc and a brush pressing on the rim. Assuming the resistance of the disc to be negligible, what will be the current through the galvanometer when the disc is rotated at $1500 \text{ rev. min}^{-1}$? If the system is frictionless, calculate the power required to maintain the motion. (C.)

19. State the laws of electromagnetic induction and describe experiments you would perform to illustrate the factors which determine the magnitude of the induced current set up in a closed circuit.

A simple electric motor has an armature of 0.1 ohm resistance. When the motor is running on a 50-volt supply the current is found to be 5 amp. Explain this and show what bearing it has on the method of starting large motors. (L.)

20. State the laws relating to the electromotive force induced in a conductor which is moving in a magnetic field.

Describe the mode of action of a simple dynamo.

Find in volts the e.m.f. induced in a straight conductor of length 20 cm, on the armature of a dynamo and 10 cm from the axis when the conductor is moving in a uniform radial field of 0.5 Wb m^{-2} and the armature is rotating at 1000 r.p.m. (L.)

chapter thirty-seven

Magnetic Fields of Current-Carrying Conductors

In the previous chapters, the induction or flux density B in a magnetic field was used to find the force on conductors and the e.m.f. induced in conductors. In this chapter we shall see how the magnitude of B is calculated. This depends on the geometry of the conductor, that is, whether it is a straight wire, or a coil, or a solenoid. The geometry also determines the pattern of the lines of force in the field.

Law of Biot and Savart

To calculate B for any shape of conductor, Biot and Savart gave a law which can now be stated as follows: The induction or flux density δB at a point P due to a small element δl of a conductor carrying a current is given by

$$\delta B \propto \frac{I \delta l \sin \alpha}{r^2}, \quad (1)$$

where r is the distance from the point P to the element and α is the angle between the element and the line joining it to P (Fig. 37.1).

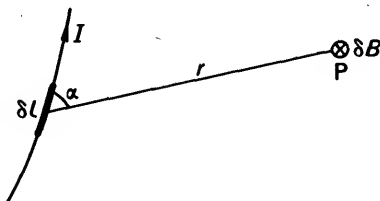


FIG. 37.1. Biot and Savart law.

The formula in (1) cannot be proved directly, as we cannot experiment with an infinitesimally small conductor. We believe in its truth because the deductions for large practical conductors turn out to be true.

The constant of proportionality in equation (1) depends on the medium in which the conductor is situated. In air (or, more exactly, in a vacuum), we write

$$\delta B = \frac{\mu_0}{4\pi} \frac{I \delta l \sin \alpha}{r^2} \quad (2)$$

The value of μ_0 is defined to be,

$$\mu_0 = 4\pi \times 10^{-7},$$

and its unit is 'henry per metre' (H m^{-1}) as will be shown later.

Induction Formula for Narrow Coil

The formula for the induction B at the centre of a narrow circular coil can be immediately deduced from (2). Here the radius r is constant for all the elements δl , and the angle α is constant and equal to 90° (Fig. 37.2 (i)). If the coil has N turns, the length of wire in it is $2\pi rN$, and the field at its centre is therefore given, if the current is I , by

$$\begin{aligned} B &= \int dB = \frac{\mu_0}{4\pi} \int_0^{2\pi rN} \frac{Idl \sin 90^\circ}{r^2} \\ &= \frac{\mu_0 I}{4\pi r^2} \int_0^{2\pi rN} dl = \frac{\mu_0 I}{4\pi r^2} 2\pi rN \\ &= \frac{\mu_0 NI}{2r} \end{aligned} \quad (1)$$

From (1), $B \propto I$ when r and N are constant, $B \propto 1/r$ when I and N are constant, and $B \propto N$ when I and r are constant. Any one of these relations may be verified with the apparatus shown in Fig. 37.2 (ii). This

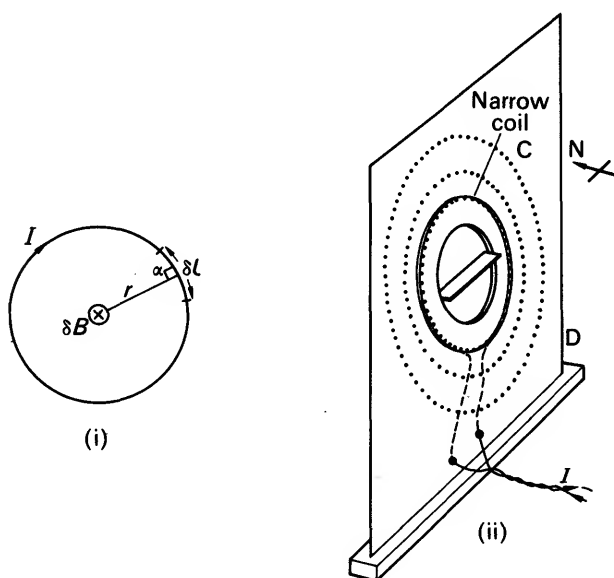


FIG. 37.2. Field of circular coil.

is a board D with sets of pegs, round which wire can be wound to form narrow coils of various radii. If a vibration magnetometer is used to measure B , the board is turned so that the axis of the coils C is in the direction of the magnetic meridian. The number of vibrations per minute is then found with the current in one direction and again when the current is reversed.

Suppose n_1 is the number when the field B of the current assists the field B_H due to the earth, so that $(B + B_H)$ is the resultant field. Suppose n_2 is the number when the field B opposes B_H and B is stronger than B_H so that $(B - B_H)$ is the resultant field. Now on p. 887, it was shown that the flux density in a field was directly proportional to the square of the frequency of the vibration magnetometer. Hence

$$B + B_H = kn_1^2 \quad \text{and} \quad B - B_H = kn_2^2.$$

Adding,

$$\therefore 2B = k(n_1^2 + n_2^2), \quad \text{or} \quad B \propto (n_1^2 + n_2^2).$$

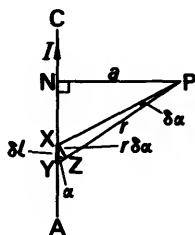
To verify $B \propto 1/r$ when I and N are constant, $(n_1^2 + n_2^2)$ is evaluated when the wire is wound round other sets of pegs and I and N are kept constant each time. A graph of $(n_1^2 + n_2^2)$ against $1/r$ is then plotted. This is found to be a straight line passing through the origin. Hence $B \propto 1/r$ when I and N are constant.

In a similar way, by varying the current I it can be shown that $B \propto I$ when N and r are constant. Similar experiments show that $B \propto N$ when I and r are constant. Thus $B \propto NI/r$ for a narrow circular coil.

The induction B in a magnetic field may also be measured by means of a *ballistic galvanometer* or by an *a.c. method*. See pp. 920, 934.

Field due to Long Straight Wire

We now deduce the induction at a point outside a long straight wire.



In Fig. 37.3 (i), AC represents part of a long straight wire. P is taken as a point so near it that, from P, the wire looks infinitely long—it subtends very nearly 180° . An element XY of this wire, of length δl , makes an angle α with the radius vector, r , from P. It therefore contributes to the magnetic field at P an amount

$$\delta B = \frac{\mu_0 I \delta l \sin \alpha}{4\pi r^2} \quad (i)$$

FIG. 37.3 (i). Field of a long, straight wire.

when the wire carries a current I .

If a is the perpendicular distance, PN, from P to the wire, then

$$PN = PX \sin \alpha$$

or

$$a = r \sin \alpha,$$

whence

$$r = \frac{a}{\sin \alpha}. \quad (ii)$$

Also, if we draw XZ perpendicular to PY, we have

$$XZ = XY \sin \alpha = \delta l \sin \alpha.$$

If δl subtends an angle $\delta \alpha$ at P, then

$$XZ = r \delta \alpha = \delta l \sin \alpha.$$

From (i),

$$\therefore \delta B = \frac{\mu_0 I \delta l \sin \alpha}{4\pi r^2} = \frac{\mu_0 I r \delta \alpha}{4\pi r^2} = \frac{\mu_0 I \delta \alpha}{4\pi r}$$

From (ii),

$$\therefore \delta B = \frac{\mu_0 I \sin \alpha \delta \alpha}{4\pi a}$$

When the point Y is at the bottom end A of the wire, $\alpha = 0$; and when Y is at the top C of the wire, $\alpha = \pi$. Therefore the total magnetic field at P is

$$B = \frac{\mu_0}{4\pi} \int_0^\pi \frac{I \sin \alpha \delta \alpha}{a} = \frac{\mu_0 I}{4\pi a} \left[-\cos \alpha \right]_0^\pi$$

$$\therefore B = \frac{\mu_0 I}{2\pi a} \quad (1)$$

Equation (1) shows that the magnetic field of a long straight wire, at a point near it, is inversely proportional to the distance of the point from the wire. The result was discovered experimentally by Biot and Savart using a vibration magnetometer method, and led to their general formula in (i) which we used to derive (1).

Variation of B with Distance—A.C. Method

An apparatus suitable for finding the variation of B with distance from a long straight wire CD is shown in Fig. 37.3 (ii). Alternating current (a.c.) of the order of 10 A, from a low voltage mains transformer, is passed through CD by using another long wire PQ at least one metre away, a rheostat B and an a.c. ammeter A. A small search coil S, with thousands of turns of wire, such as the coil from an output transformer, is placed near CD. It is positioned with its axis at a small distance r from CD

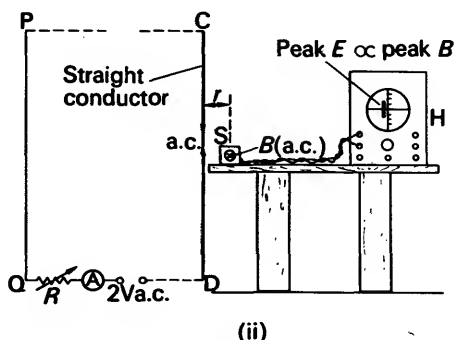


FIG. 37.3 (ii). Investigation of B due to long straight conductor.

and so that the flux from CD enters its face normally. S is joined by long twin flex to the Y-plates of an oscilloscope H and the greatest sensitivity, such as 5 mV/cm, is used.

When the a.c. supply is switched on, the varying flux through S produces an induced alternating e.m.f. E . The peak value of E can be determined by switching off the time-base and measuring the length of the line trace, Fig. 37.3 (ii). See p. 1014. Now the peak value of B , the magnetic induction, is proportional to the peak value of E , as shown in the case of the simple dynamo on p. 907. Thus the length of the trace gives a measure of the peak value of B .

The distance r of the coil from CD is then increased and the corresponding length of the trace is measured. The length of the trace plotted against $1/r$ gives a straight line graph passing through the origin. Hence $B \propto 1/r$. A similar method can be used for investigating the induction B for the case of a narrow circular coil or for a solenoid (p. 939).

EXAMPLE

Calculate the flux density at a distance of 1 cm or 0.01 m from a very long vertical straight wire carrying a current of 10 A. At what distance from the wire will the field induction neutralize that due to the earth's horizontal component flux-density, 0.2×10^{-4} T?

$$(i) \quad B = \frac{\mu_0 I}{2\pi a} = \frac{4\pi \times 10^{-7} \times 10}{2\pi \times 10^{-2}} \\ = 2 \times 10^{-4} \text{ T.}$$

(ii) 1 cm from the wire, the induction due to the current is 2×10^{-4} T.

Now B is inversely-proportional to a , the distance from the wire. Thus B is 0.2×10^{-4} Wb m^{-2} , or ten times smaller than at 1 cm, at a distance 10 times as great. Thus the distance is 10 cm.

Note that the actual position of the point where the two fields neutralize must take account of the fact that B is a *vector*, that is, it has direction and magnitude. For a downward current of 10 A, the point concerned is due east of the wire. It is called a *neutral* point.

Field along Axis of a Narrow Circular Coil

We will now find the magnetic field at a point anywhere on the axis of a narrow circular coil (P in Fig. 37.4). We consider an element δl of the coil, at right angles to the plane of the paper. This sets up a field δB at P, in the plane of the paper, and at right angles to the radius vector r . If β is the angle between r and the axis of the coil, then the field δB has components $\delta B \sin \beta$ along the axis, and

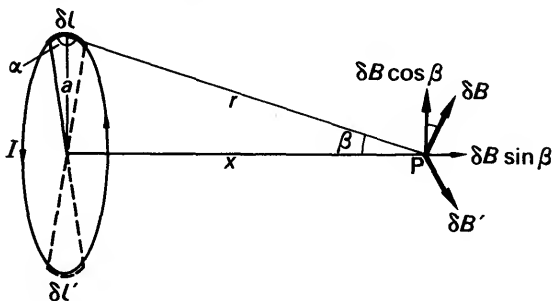


FIG. 37.4. Field on axis of flat coil.

$\delta B \cos \beta$ at right angles to the axis. If we now consider the element $\delta l'$ diametrically opposite to δl , we see that it sets up a field $\delta B'$ equal in magnitude to δB . This also has a component, $\delta B' \cos \beta$, at right angles to the axis; but this component acts in the opposite direction to $\delta B \cos \beta$ and therefore cancels it. By considering elements such as δl and $\delta l'$ all round the circumference of the coil, we see that the field at P can have no component at right angles to the axis. Its value along the axis is

$$B = \int dB \sin \beta.$$

From Fig. 37.4, we see that the length of the radius vector r is the same for all points on the circumference of the coil, and that the angle α is also constant, being 90° . This, if the coil has a single turn, and carries a current I ,

$$\delta B = \frac{\mu_0 I \delta l \sin \alpha}{4\pi r^2} = \frac{\mu_0 I}{4\pi r^2} \delta l.$$

And, if the coil has a radius a , then

$$\begin{aligned} B &= \int dB \sin \beta = \int_0^{2\pi a} \frac{\mu_0 I}{4\pi r^2} dl \sin \beta \\ &= \frac{\mu_0 I a \sin \beta}{2r^2}. \end{aligned} \quad (i)$$

When the coil has more than one turn, the distance r varies slightly from one turn to the next. But if the width of the coil is small compared with all its other dimensions, we may neglect it, and write,

$$B = \frac{\mu_0 N I a \sin \beta}{2r^2}, \quad (ii)$$

where N is the number of turns.

Equation (i) can be put into a variety of forms, by using the facts that

$$\sin \beta = \frac{a}{r},$$

and

$$r^2 = x^2 + a^2,$$

where x is the distance from P to the centre of the coil. Thus

$$B = \frac{\mu_0 N I a^2}{2r^3} = \frac{\mu_0 N A I}{2\pi(x^2 + a^2)^{3/2}}, \quad (1)$$

When the distance x is large compared with a , the expression (1) reduced to:

$$B = \frac{\mu_0 N A I}{2\pi x^3} = \frac{\mu_0 m}{2\pi x^3}, \quad (2)$$

where $m = N A I$ = the *magnetic moment* of the circular coil, p. 885.

Helmholtz Coils

The field along the axis of a single coil varies with the distance x from the coil. In order to obtain a *uniform* field, Helmholtz used two coaxial parallel coils of equal radius R , separated by a distance R . In this case, when the same current flows round each coil in the same direction, the resultant field B is uniform for some distance on either side of the point on their axis midway between the coils. This may be seen roughly by adding the fields due to each coil alone. Helmholtz coils were used in Thomson's determination of e/m (p. 1003).

The magnitude of the resultant field B at the midpoint can be found from our previous formula for a single coil. We now have $a = R$ and $x = R/2$. Thus, for the two coils,

$$\begin{aligned}
 B &= 2 \times \frac{\mu_0 N I R^2}{2(R^2/4 + R^2)^{3/2}} = \left(\frac{4}{5}\right)^{3/2} \times \frac{\mu_0 N I}{R} \\
 &= 0.72 \frac{\mu_0 N I}{R} \text{ (approx.).}
 \end{aligned}$$

Field on Axis of a Long Solenoid

We may regard a solenoid as a long succession of narrow coils; if it has n turns per metre, then in an element δx of it there are $n\delta x$ coils (Fig. 37.5). At a point P on the axis of the solenoid, the field due to these is, by equation (ii),

$$\delta B = \frac{\mu_0 I a \sin \beta}{2r^2} n \delta x,$$

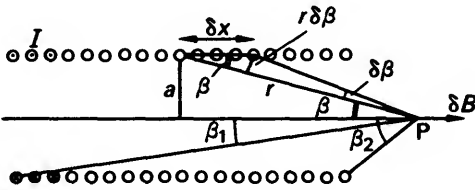


FIG. 37.5. Field on axis of solenoid.

in the notation which we have used for the flat coil. If the element δx subtends an angle $\delta\beta$ at P , then, from the figure,

$$r\delta\beta = \delta x \sin \beta;$$

whence

$$\delta x = \frac{r\delta\beta}{\sin \beta}$$

Also,

$$a = r \sin \beta.$$

Thus

$$\begin{aligned}
 \delta B &= \frac{\mu_0 I r \sin^2 \beta}{2r^2} n \frac{r\delta\beta}{\sin \beta} \\
 &= \frac{\mu_0 n I}{2} \sin \beta \delta\beta.
 \end{aligned}$$

If the radii of the coil, at its ends, subtend the angles β_1 and β_2 at P , then the field at P is

$$\begin{aligned}
 H &= \int_{\beta_1}^{\beta_2} \frac{\mu_0 n I}{2} \sin \beta d\beta \\
 &= \frac{\mu_0 n I}{2} \left[-\cos \beta \right]_{\beta_1}^{\beta_2}, \\
 &= \frac{\mu_0 n I}{2} (\cos \beta_1 - \cos \beta_2). \quad \quad \quad (1)
 \end{aligned}$$

The magnetic field within a toroid is very nearly uniform, because the coil has no ends. The coil is equivalent to an infinitely long solenoid, and the field-strength at all points within it is given by

$$B = \mu_0 n I \quad (2)$$

A 'Slinky' is a coil which can be stretched to provide simply a solenoid with a varying number of turns per metre, n . A small search coil with many thousands of turns, placed coaxially inside the solenoid, can be connected to an oscilloscope to provide a measure of B when alternating current is passed into the solenoid. See p. 934. Since $n \propto 1/L$, where L is the length of the coil, a graph of B against $1/L$ can be plotted for various values of L , the current being the same each time. A straight line through the origin is obtained, showing that $B \propto n$. A search coil connected to a ballistic galvanometer, and a direct current which is reversed in the coil, may also be used to provide a measure of B (p. 919).

Forces between Currents

Ampère carried out many experiments on the forces of attraction and repulsion between two current-carrying conductors. Each was acted on by the field of the other, as shown in Fig. 37.8. Currents flowing in the same direction ('like' currents) attracted each other,

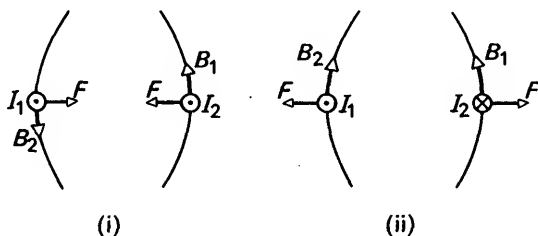


FIG. 37.8. Forces between currents.

Fig. 37.8 (i), while 'unlike' currents (in opposite directions) repelled each other, Fig. 37.8 (ii). This difference from the laws governing poles and charges greatly impressed Ampère.

If two long straight conductors lie parallel and close together at a distance r apart, and carry currents I, I' respectively, then the current I is in a magnetic field of flux density B equal to $\mu_0 I' / 2\pi r$ due to the current I' (p. 934). The force per metre length F is hence given by

$$F = B I l = B I \times 1 = \frac{\mu_0 I'}{2\pi r} \times I \times 1$$

$$\therefore F = \frac{\mu_0 I I'}{2\pi r} \quad (1)$$

Nowadays the ampere is *defined* in terms of the force between conductors. It is *that current, which flowing in each of two infinitely-long*

parallel straight wires of negligible cross-sectional area separated by a distance of 1 metre in vacuo, produces a force between the wires of 2×10^{-7} newton metre $^{-1}$.

Taking $I = I' = 1$ A, $r = 1$ metre, $F = 2 \times 10^{-7}$ newton metre $^{-1}$, then, from (1),

$$2 \times 10^{-7} = \frac{\mu_0 \times 1 \times 1}{2\pi \times 1}$$

$$\therefore \mu_0 = 4\pi \times 10^{-7} \text{ henry metre}^{-1},$$

which is the value quoted above.

Unit of μ_0

The permeability of free space μ_0 was defined previously from the relation

$$\delta B = \frac{\mu_0 I \delta s \sin \theta}{4\pi r^2}$$

From this relation, the unit of μ_0 is

$$\frac{\text{weber metre}^{-2} \times \text{metre}^2}{\text{ampere} \times \text{metre}} \text{ or } \text{Wb A}^{-1} \text{ m}^{-1} \quad (2)$$

Now the unit of inductance L is the henry (H), which can be defined from the relation $\Phi = LI$ (p. 925). Thus, since $L = \Phi/I$

$$1 \text{ H} = 1 \text{ Wb A}^{-1}.$$

From (2), it follows that the unit of μ_0 can be written as

$$\text{H m}^{-1} \text{ (henry per metre),}$$

and this is the SI unit of μ_0 and of permeability μ generally.

Absolute Determination of Current

A laboratory form of an *ampere balance*, which measures current by measuring the force between current-carrying conductors, is shown in Fig. 37.9.

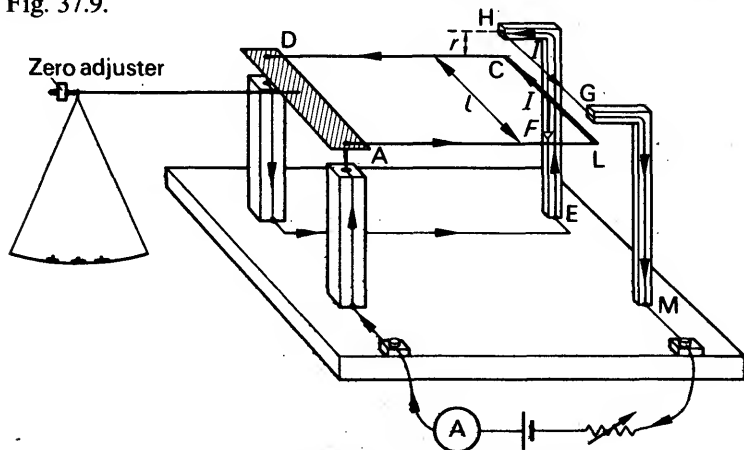


FIG. 37.9. Laboratory form of Ampere Balance.

With no current flowing, the zero screw is adjusted until the plane of ALCD is horizontal. The current I to be measured is then switched on so that it flows through ALCD and EHGM in series and HG repels CL. The mass m necessary to restore balance is then measured, and mg is the force between the conductors since the respective distances of CL and the scale pan from the pivot are equal. The equal lengths l of the straight wires CL and HG, and their separation r , are all measured.

From equation (1) on p. 939,

$$\text{force per metre} = \frac{4\pi \times 10^{-7} I^2}{2\pi r}$$

$$\therefore mg = \frac{4\pi \times 10^{-7} I^2 l}{2\pi r}$$

$$\therefore I = \sqrt{\frac{mgr}{2 \times 10^{-7} l}}$$

In this expression, I will be in amperes if m is in kilogrammes, $g = 9.8 \text{ m s}^{-2}$ and l and r are measured in metres.

Magnetizing Force, or Intensity, H

Biot and Savart's law has been stated as

$$\delta B = \frac{\mu_0 I \delta l \sin \alpha}{4\pi r^2}$$

This is only true in air or a vacuum. In other materials the flux density may be altered, even though the currents remain the same. To take account of this we write

$$\delta B = \frac{\mu I \delta l \sin \alpha}{4\pi r^2}$$

where μ is the *permeability* of the medium. μ_0 is called the permeability of 'free space' or vacuum.

So far we have only used the flux density, B , in a field. Another field quantity, symbol H , is also used. It is called the *magnetizing force or intensity*, and is defined by the relation

$$H = \frac{B}{\mu}$$

We may thus write δH arising from a current I in an element of length δl as

$$\delta H = \frac{I \delta l \sin \alpha}{4\pi r^2} \quad \dots \quad (1)$$

From this it can be seen that the unit of H is

$$\frac{\text{ampere metre}}{\text{metre}^2} \quad \text{or} \quad \text{ampere metre}^{-1} (\text{A m}^{-1});$$

whereas the unit of B in the tesla (T) or weber metre⁻² (Wb m⁻²).

In any medium δB has a value depending on the permeability of the medium. From (1), it can be seen that H does not depend on μ but

only the currents and their geometry. H is *independent*, therefore, of the medium in which the conductors are situated. It is for this reason that H is regarded as being due directly to the currents. H is then a 'cause' which gives rise to a flux density B given by μH , and so B is dependent on the medium used. Because of this interpretation, H is often called the 'magnetizing force', or 'magnetizing intensity'.

From equation (1) on p. 934, the magnetizing force due to a straight wire is given by

$$H = \frac{I}{2\pi a}.$$

Ampère's Theorem

In the calculation of magnetic fields, we have used so far only the Biot and Savart law. Another law useful for calculating magnetic field strengths is *Ampère's theorem*.

Consider Fig. 37.10, in which a continuous closed line or loop L is drawn round the wires P , Q , R which carry currents of I_1 , I_2 , I_3 respectively. The total current enclosed by L is $(I_1 + I_2 + I_3)$. Now if H is the magnetizing force or *magnetic field intensity* at any element dl

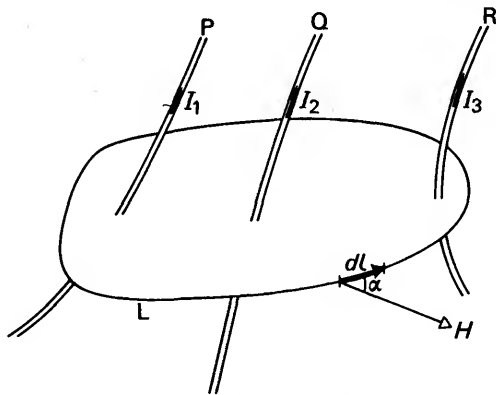


FIG. 37.10. Ampère's theorem.

in the loop L , obtained from the Biot and Savart law, it is possible to calculate the quantity $H \cdot dl \cdot \cos \alpha$, where α is the angle between the element and the field H . When the quantity $H \cdot dl \cdot \cos \alpha$ is summed for every element in the closed loop L , calculation beyond the scope of this book shows that the result is $(I_1 + I_2 + I_3)$. This is expressed by

$$\oint H \cdot dl \cdot \cos \alpha = I_1 + I_2 + I_3,$$

where the symbol \oint represents the integral taken completely round the closed loop. Ampère's theorem is the general statement

$$\oint H \cdot dl \cdot \cos \alpha = I,$$

where I is the total current enclosed by the loop.

We now apply the theorem to two special cases of current-carrying conductors.

1. *Straight wire*

Fig. 37.11 shows a circular loop L of radius r , drawn concentrically round a straight wire carrying a current I . The lines of force are circles and hence, at every part of a closed line, H is directed along the line itself. Thus $\alpha = 0^\circ$ all round the line. Further, by symmetry, H has the same value everywhere on the line.

$$\therefore \oint H \cdot dl \cdot \cos \alpha = H \oint dl = H \cdot 2\pi r,$$

since $\alpha = 0^\circ$ and H is constant. Hence, from Ampère's theorem,

$$H \cdot 2\pi r = I$$

$$\therefore H = \frac{I}{2\pi r}, \quad \text{and} \quad B = \mu_0 H = \frac{\mu_0 I}{2\pi r}.$$

This agrees with the result for B on p. 934.

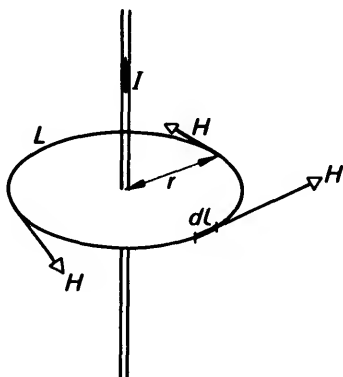


FIG. 37.11. Field intensity of straight wire.

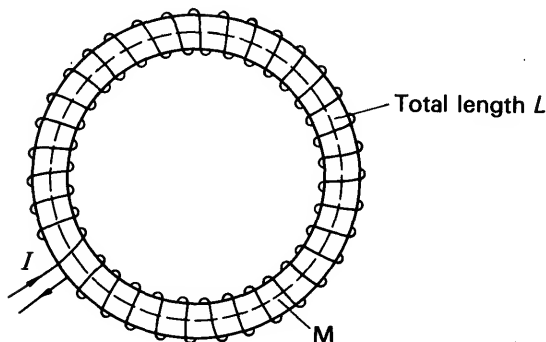


FIG. 37.12. Field intensity of toroid.

2. *Toroid*

Consider the closed loop M indicated by the broken line in Fig. 37.12. Again H is everywhere the same on M and is directed along the loop.

$$\begin{aligned} \therefore \oint H \cdot dl \cos \alpha &= H \oint dl \\ &= HL, \end{aligned}$$

where L is the total length of the loop M . Hence, from Ampere's theorem,

$$HL = NI$$

$$\therefore H = \frac{NI}{L} = nI,$$

where N is the total number of turns, and n is the number of turns per metre. This agrees with p. 938.

Earth's Magnetism

It was Dr. Gilbert who first showed that a magnetized needle, when freely suspended about its centre of gravity, dipped downwards towards the north at about 70° to the horizontal in England. He also found that this *angle of dip* increased with latitude, as shown in Fig. 37.13, and concluded that the earth itself was, or contained, a magnet. The points where the angle of dip is 90° are called the earth's magnetic poles; they

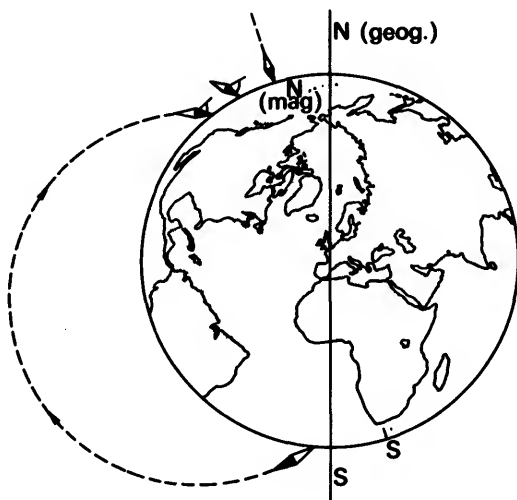


FIG. 37.13. Illustrating the angle of dip.

are fairly near to the geographic poles, but their positions are continuously, though slowly, changing. Gilbert's simple idea of the earth as a magnet has had to be rejected. The earth's crust does not contain enough magnetic material to make a magnet of the required strength; the earth's core is, we believe, molten—and molten iron is non-magnetic. The origin of the earth's magnetism is, in fact, one of the great theoretical problems of the present day.

Horizontal and Vertical Components. Variation and Dip

Since a freely suspended magnetic needle dips downward at some angle δ to the horizontal, the earth's *resultant magnetic field*, B_R acts at an angle δ to the horizontal. The 'angle of dip', or *inclination*, can

thus be defined as the angle between the resultant earth's field and the horizontal. The earth's field has a *vertical component*, B_v , given by

$$B_V = B_R \sin \delta, \quad (1)$$

and a horizontal component, B_H , given by

$$B_H = B_R \cos \delta. \quad (2)$$

Also, $\frac{B_v}{B_H} = \tan \delta$. (3)

To specify the earth's magnetic field at any point, we must state its

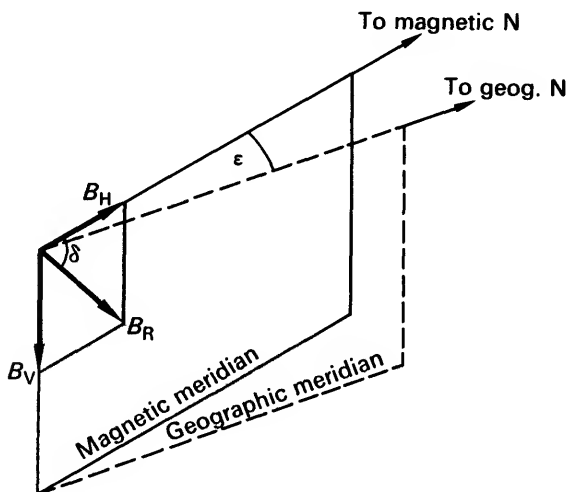


FIG. 37.14. Magnetic and geographic meridians. Dip.

strength and direction. To specify its direction we must give the direction of the magnetic meridian, and the angle of dip δ (Fig. 37.14). In most parts of the world the magnetic meridian does not lie along the geographic meridian (the vertical plane running geographically north-south). The angle between the magnetic and geographic meridians, ϵ , is called the magnetic variation, or sometimes the declination, at the place concerned; it is shown on the margins of maps. The horizontal and vertical components of the earth's field, and the angle of dip, can be measured by a large coil or 'earth inductor' (see p. 921).

EXERCISES 37

1. Define the *ampere*. Write down expressions for (i) the magnetic field strength (magnetizing force) at a distance of d from a very long straight conductor carrying a current I , and (ii) the mechanical force acting on a straight conductor of length l carrying a current I at right angles to a uniform magnetic field of flux density B .

Show how these two expressions may be used to deduce a formula for the force per unit length between two long straight parallel conductors in vacuo carrying currents I_1 and I_2 separated by a distance d .

A horizontal straight wire 5 cm long weighing 1.2 g m^{-1} is placed perpendicular to a uniform horizontal magnetic field of flux density 0.6 Wb m^{-2} . If the resistance of the wire is 3.8 ohm m^{-1} , calculate the p.d. that has to be applied between the ends of the wire to make it just self-supporting. Draw a diagram showing the direction of the field and the direction in which the current would have to flow in the wire. (C.)

2. State the law of force acting on a conductor carrying an electric current in a magnetic field. Indicate the direction of the force and show how its magnitude depends on the angle between the conductor and the direction of the field.

Sketch the magnetic field due solely to two long parallel conductors carrying respectively currents of 12 and 8 A in the same direction. If the wires are 10 cm apart, find where a third parallel wire also carrying a current must be placed so that the force experienced by it shall be zero. (L.)

3. Define the *ampere*.

Two long vertical wires, set in a plane at right angles to the magnetic meridian, carry equal currents flowing in opposite directions. Draw a diagram showing the pattern, in a horizontal plane, of the magnetic flux due to the currents alone—that is, for the moment ignoring the earth's magnetic field.

Next, taking into account the earth's magnetic field, discuss the various situations that can give rise to neutral points in the plane of the diagram.

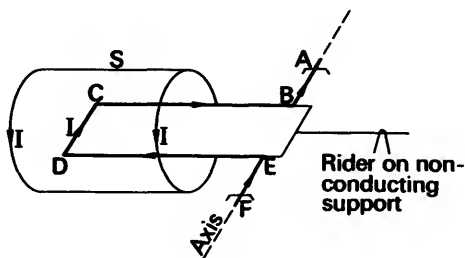


FIG. 37.15.

Fig. 37.15 shows a simple form of current balance. The 'long' solenoid S , which has 2000 turns per metre, is in series with the horizontal rectangular copper loop $ABCDEF$, where $BC = 10 \text{ cm}$ and $CD = 3 \text{ cm}$. The loop, which is freely pivoted on the axis AF , goes well inside the solenoid, and CD is perpendicular to the axis of the solenoid. When the current is switched on, a rider of mass 0.2 g placed 5 cm from the axis is needed to restore equilibrium. Calculate the value of the current, I . (O.)

4. Define *magnetic moment*.

A small magnet, suspended with its axis horizontal so as to be able to rotate freely about a vertical axis, is situated at the centre of a long horizontal solenoid,

the axis of which lies at right angles to the magnetic meridian. If the solenoid has 20 turns per cm, determine the value of the current passing through it which would cause the magnet to rotate through 50° . (Horizontal component of earth's magnetic field intensity = 14 A m^{-1} .) (N.)

5. Describe with experimental details how you would carry out any *two* of the following in the laboratory:

- Determine the current sensitivity of a moving-coil galvanometer.
- Determine the internal resistance of a dry cell using a potentiometer.
- Investigate, using a vibration magnetometer, how the magnetic field due to a long straight wire carrying a steady current varies with distance from the wire. (L.)

6. Define the *ampere*.

Draw a labelled diagram of an instrument suitable for measuring a current absolutely in terms of the ampere, and describe the principle of it.

A very long straight wire PQ of negligible diameter carries a steady current I_1 . A square coil ABCD of side l with n turns of wire also of negligible diameter is set up with sides AB and DC parallel to and coplanar with PQ; the side AB is nearest to PQ and is at a distance d from it. Derive an expression for the resultant force on the coil when a steady current I_2 flows in it, and indicate on a diagram the direction of this force when the current flows in the same direction in PQ and AB.

Calculate the magnitude of the force when $I_1 = 5 \text{ A}$, $I_2 = 3 \text{ A}$, $d = 3 \text{ cm}$, $n = 48$ and $l = 5 \text{ cm}$. (O. & C.)

7. (i) Draw a diagram showing the pattern of the magnetic field lines in a horizontal plane passing through the centre of a plane vertical circular coil carrying a steady current. Indicate on your diagram possible positions for neutral points if the plane of the coil were to be set in the magnetic meridian.

(ii) How would you show experimentally that the value of B at the centre of such a coil is proportional to nl/r , where n is the number of turns and r the radius?

8. A long straight wire carries a steady current I . Write down a formula for the magnetic intensity (magnetizing force) H due to the current at a point distant y from the wire. Give a consistent set of units for the quantities in your formula.

Show on a diagram the lines of magnetic force due to the currents when two parallel wires separated by a distance $2a$ carry equal currents in the same direction. Use your formula to derive an expression for the magnetic intensity at a point P in a plane perpendicular to the wires at a distance x from the point midway between the wires and along the right bisector of the line joining them in this plane. Hence derive the condition that the intensity at P shall be a maximum. (N.)

chapter thirty-eight

Magnetic Properties of Materials

THE magnetic properties of materials require investigation to decide whether they are suitable for permanent magnets such as loudspeaker magnets, for temporary magnets such as electromagnets, or for cores of electromagnetic induction machines such as transformers.

Induction or Flux Density in Magnetic Material

Consider a toroid of length L , wound with N turns each carrying a current I round a ring of magnetic material, Fig. 38.1.

The *total flux density* B in the material is partly due to the currents

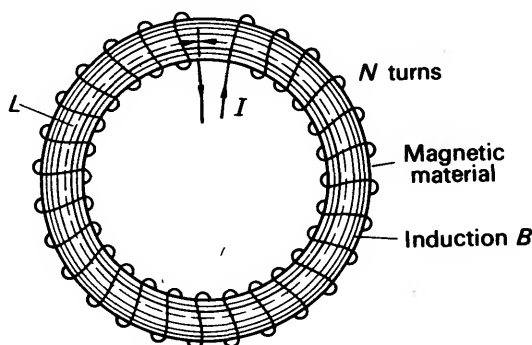


FIG. 38.1. Induction in magnetic material.

flowing in the wire and partly due to the magnetization of the material. We thus write

$$B = B_0 + B_M, \quad (1)$$

where B_0 is the flux density due directly to the current in the wire, and B_M is the flux density due to the magnetization of the material.

We now assume that the induction B_M is produced by many small circulating currents inside the magnetic material, due to the circulating and spinning electrons in the atoms. Fig. 38.2 shows that the effect of many small adjacent current loops may be thought of as one current loop. In the same way, the internal circulating and spinning currents can be replaced by a single current I_M flowing in the coil wound round

the core. This theoretical surface or magnetization current is additional to the real or actual current I flowing in the coil.

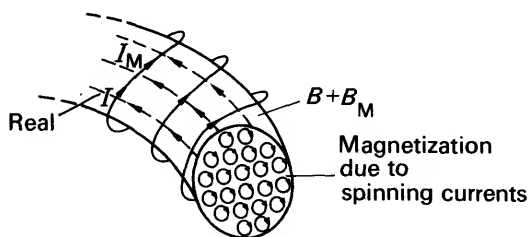


FIG. 38.2. Magnetizing surface current.

By itself, the real current I produces a flux-density B_0 . If the toroid has n turns per unit length ($n = N/L$), then, from p. 938,

$$B_0 = \mu_0 n I.$$

The surface current I_M may be imagined to flow in n turns per metre of the solenoid, as the real current does. The flux density or induction B_M is then given by

$$B_M = \mu_0 n I_M.$$

This is the induction which would be produced by the current I_M if the material were not present in the toroid, that is, a current I_M in the coil would produce a flux density equal to that due to the magnetization in the material.

$$\therefore \text{total induction } B = B_0 + B_M = \mu_0 n (I + I_M) \quad (2)$$

Intensity of Magnetization

It is possible to write B_M , the induction due to the magnetization of the material, in a different way. The magnetic moment of each turn due to this imaginary surface current $= A \times I_M$, where A is the area of each turn. See p. 885. The magnetic moment of the whole toroid is then $nLA I_M$, since nL is the total number of turns. Hence the *magnetic moment per unit volume*

$$= \frac{nLA I_M}{\text{volume}} = \frac{nLA I_M}{LA} = n I_M$$

The 'magnetic moment per unit volume' is called the *intensity of magnetization*, M , of the magnetic core in the toroid. Thus $B_M = \mu_0 n I_M = \mu_0 M$. Further, nI is the magnetic field intensity, H , in the toroid due to the current I .

Thus the *total induction or flux-density* B in the core when the coil carries a current is given by

$$B = B_0 + B_M = \mu_0 n I + \mu_0 n I_M$$

$$\therefore B = \mu_0 (H + M) \quad (3)$$

Equation (3) relates the total induction B to the magnetizing field intensity H due to the current and to the intensity of magnetization M of the material produced. Note that H and M have the same units from (3). Thus the unit of M is ampere per metre (A m^{-1}).

EXAMPLE

A solenoid 1 m long is wound with 10^4 turns of copper wire each carrying a current of 10 A. If the cross-sectional area of the coil is 10 cm^2 , calculate (a) the magnetizing force or intensity in the solenoid, (b) the couple exerted on the solenoid when it is placed at right angles to an external field of induction 10^{-2} T .

If the solenoid is now filled with certain magnetic material, the induction in the material is 1.5 T. Calculate (c) the intensity of magnetization in the material, (d) the total couple exerted when the solenoid and material inside is placed at right angles to a field of induction 10^{-2} T .

- (a) We have $H = nI$, where n is the number of turns per metre,
 $= 10^4 \times 10$
 $= 10^5 \text{ ampere metre}^{-1} (\text{A m}^{-1})$.

- (b) Magnetic moment of each solenoid turn $= IA$.

$$\therefore \text{magnetic moment of empty solenoid, } m, = NIA.$$

Since $N = 10^4$, $I = 10 \text{ A}$, $A = 10 \text{ cm}^2 = 10 \times 10^{-4} \text{ m}^2$,

$$\begin{aligned}\therefore m &= 10^4 \times 10 \times 10 \times 10^{-4} \\ &= 100 \text{ A m}^2.\end{aligned}$$

Now couple $C = mB \sin \alpha$, where α is the angle between the solenoid axis and the external field B .

$$\begin{aligned}\therefore C &= mB \sin 90^\circ \\ &= 100 \times 10^{-2} \\ &= 1 \text{ newton metre (N m)}\end{aligned}$$

- (c) We have $B = \mu_0(H + M)$

$$\begin{aligned}\therefore M &= \frac{B - \mu_0 H}{\mu_0} \\ &= \frac{1.5 - 4\pi \times 10^{-7} \times 10^5}{4\pi \times 10^{-7}} \\ &= 10.9 \times 10^5 \text{ ampere metre}^{-1}\end{aligned}$$

- (d) Total magnetic moment, m , of the specimen is given by,

$$\begin{aligned}m &= MV \\ &= 10.9 \times 10^5 \times (1 \times 10^{-3}) \\ &= 10.9 \times 10^2 \text{ A m}^2\end{aligned}$$

\therefore Total magnetic moment of the solenoid and the material inside

$$\begin{aligned}&= 100 + 10.9 \times 10^2 \\ &= 11.9 \times 10^2 \text{ A m}^2\end{aligned}$$

$$\begin{aligned}
 \therefore \text{Couple} &= mB \sin \alpha \text{ newton metre} \\
 &= 11.9 \times 10^2 \times 10^{-2} \times \sin 90^\circ \\
 &= 11.9 \text{ N m.}
 \end{aligned}$$

Relative Permeability

The permeability of a material is defined by the relation $\mu = B/H$. Now, in general, $B = \mu_0(H + M)$

$$\therefore \mu H = \mu_0(H + M)$$

$$\text{or} \quad \frac{\mu}{\mu_0} = 1 + \frac{M}{H}$$

The ratio μ/μ_0 is called the *relative permeability*, μ_r , of the material. The ratio M/H is called the *susceptibility*, χ , of the material. Hence, from above,

$$\mu_r = 1 + \chi. \quad (4)$$

It should be noted that μ_r and χ are dimensionless; each is the ratio of two quantities having the same dimensions. Permeability, μ , however, which is the ratio B/H , has dimensions; its unit is *henry per metre* (H m^{-1}). See p. 940.

We shall describe shortly how the variation of B with H may be measured. Here we shall anticipate the results. As a magnetic material, originally unmagnetized, is subjected to an increasing field, the intensity of magnetization M increases until it reaches a maximum value (Fig. 38.3 (a)). The material is then 'saturated'; that is, its magnetic

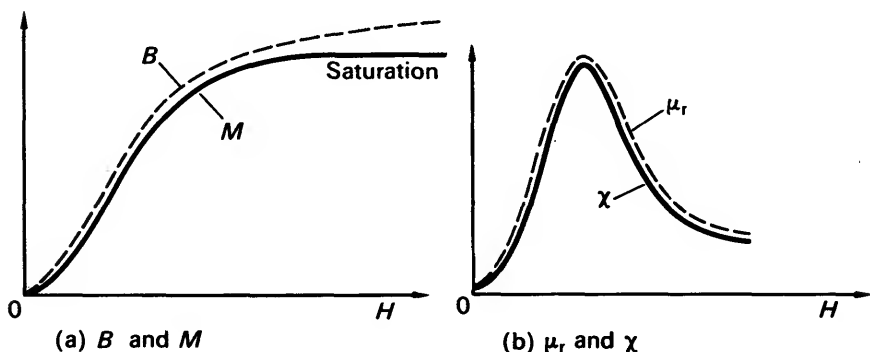


FIG. 38.3. Variation of B , M , μ_r , χ .

'domains' are completely aligned with the field H . B , however, continues to increase with H , since $B = \mu_0(H + M)$. Fig. 38.3 (b) also shows how the relative permeability μ_r and the susceptibility χ varies with H . It increases at first, and passes through a maximum value. As the material approaches saturation the domains cannot yield much further, and the susceptibility falls to a low value.

Variation of B with H

Ferromagnetic materials are those which have a high susceptibility. This is generally very much greater than 1, for example, 3000. Some materials are capable of retaining their magnetization and forming strong permanent magnets. Others form temporary magnets (p. 954).

The relationship between B , the flux density in a material, and the applied magnetizing field or force H , is best investigated experimentally by using a toroid shaped specimen. This eliminates the reduction in the field due to the effect of poles at the ends of a cylindrical rod. However, in the experiment described shortly, a specimen is placed inside a long current-carrying solenoid. This does not produce as large a value B as that obtained with a toroid, but the essential features of the variation of B with H are still observed.

One form of apparatus is shown in Fig. 38.4. The current I through the solenoid is measured on the ammeter A , and since $H = nI$, the current I is directly proportional to the magnetizing force H . C is a

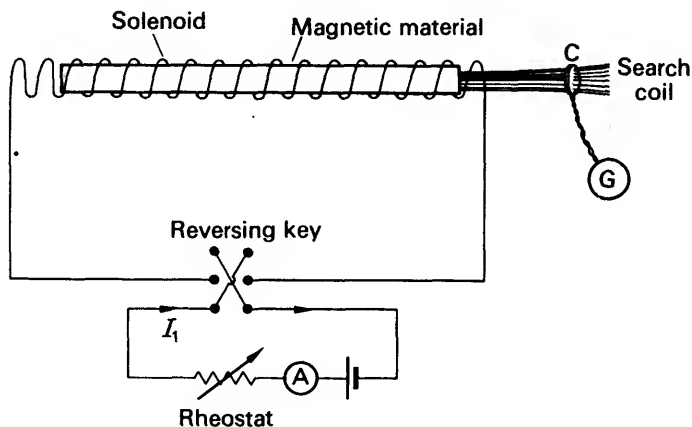


FIG. 38.4. Variation of B with H by experiment.

small search coil, which is placed in contact with the end of the specimen, and which is connected to a ballistic galvanometer G . When the coil is sharply removed from the vicinity of the specimen the throw on the galvanometer gives a measure of the flux change (see p. 920). Since the number of turns and area of the coil is constant, this throw will be proportional to the flux density B .

This specimen is first demagnetized (see p. 958) and placed in the solenoid. The current I is now increased in steps from zero, and the corresponding deflections θ on the ballistic galvanometer are observed. The specimen is taken through a *magnetic cycle* of magnetization. This is done by increasing I until θ is nearly constant, when the specimen has become saturated, then reducing I to zero and reversing it until saturation is reached in the opposite direction, and finally reducing I to zero and increasing it once more in the opposite direction.

H can be calculated from the relation $H = nI$. If the search coil has N turns of area A , and the ballistic galvanometer is calibrated, then B can be found from:

$$\text{Charge} = \frac{\text{Change in flux}}{\text{Resistance of circuit, } R}$$

or
$$Q = \frac{BAN}{R}$$

If R and Q are measured, then B can be found.

Sometimes it is required to plot M against H rather than B against H . In this case M is calculated, using the relation (3) on p. 949, from

$$M = \frac{B - \mu_0 H}{\mu_0}$$

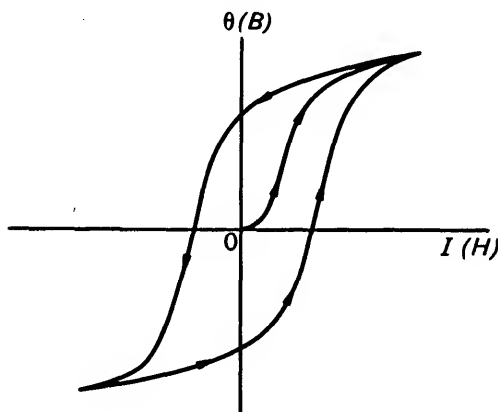


FIG. 38.5. B - H variation.

If only the general form of the B - H curve is wanted, it is sufficient to plot θ against I . Fig. 38.5 illustrates a typical graph obtained; the arrows show the sequence in going round the magnetic cycle.

Hysteresis. Remanence. Coercive Force

Fig. 38.6 shows the variation of magnetic induction, B , with the applied field, H , when the specimen is taken through a complete cycle. After the specimen has become saturated, and the field is reduced to zero, the iron is still quite strongly magnetized, setting up a flux-density B_r . This flux-density is called the *remanence*; it is due to the tendency of groups of molecules, or domains, to stay put once they have been aligned.

When the field is reversed, the residual magnetism is opposed. Each increase of magnetizing field now causes a decrease of flux-density, as the domains are twisted farther out of alignment. Eventually, the flux-density is reduced to zero, when the opposing field H has the value

H_c . This value of H is called the *coercive force* of the iron; it is a measure of the difficulty of breaking up the alignment of the domains.

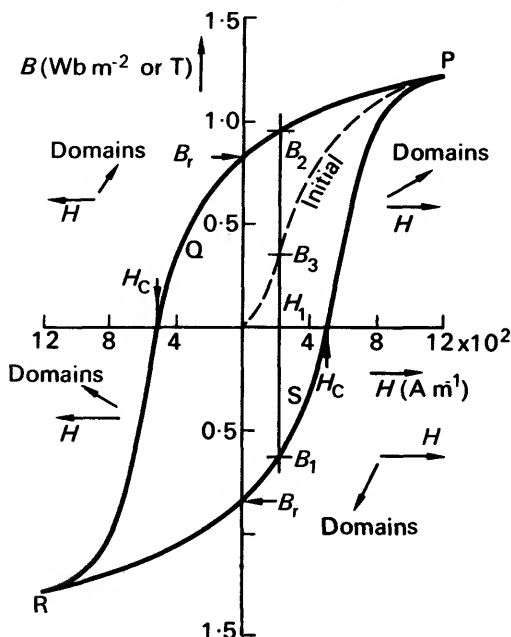


FIG. 38.6. Hysteresis loop.

We now see that, when once the iron has been magnetized, its magnetization curve never passes through the origin again. Instead, it forms the closed loop PQRS, which is called a *hysteresis loop*. Hysteresis, which comes from a Greek work meaning 'delayed', can be defined as *the lagging of the magnetic induction, B , behind the magnetizing field, H , when the specimen is taken through a magnetic cycle.*

Properties of Magnetic Materials

Fig 38.7 shows the hysteresis loops of iron and steel. Steel is more suitable for permanent magnets, because its high coercivity means that it is not easily demagnetized by shaking. The fact that the remanence of iron is a little greater than that of steel is completely outweighed by its much smaller coercivity, which makes it very easy to demagnetize. On the other hand, iron is much more suitable for electromagnets, which have to be switched on and off, as in relays. Iron is also more suitable for the cores of transformers and the armatures of machines. Both of these go through complete magnetizing cycles continuously: transformer cores because they are magnetized by alternating current, armatures because they are turning round and round in a constant field. In each cycle the iron passes through two parts of its hysteresis loop (near Q and S in Fig. 38.6), where the magnetizing field is having to demagnetize the iron. There the field is doing work against the internal

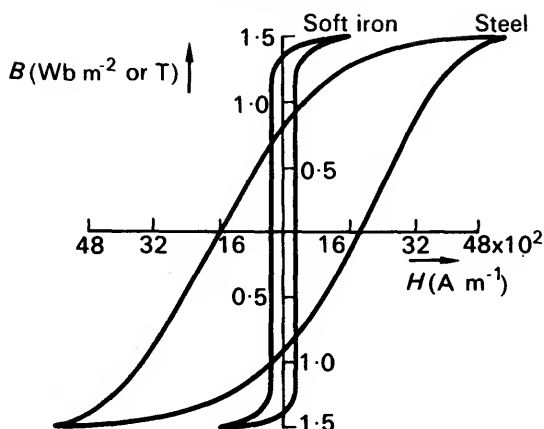


FIG. 38.7. Hysteresis curves for iron and steel.

friction of the domains. This work, like all work that is done against friction, is dissipated as heat. The energy dissipated in this way, per cycle, is less for iron than for steel, because iron is easier to demagnetize. It is called the *hysteresis loss*; we will show soon that it is proportional to the area of the B - H loop (p. 957).

In a large transformer the hysteresis loss, together with the heat developed by the current in the resistance of the windings, liberates so much heat that the transformer must be artificially cooled. The cooling is done by circulating oil, which itself is cooled by the atmosphere: it passes through pipes which can be seen outside the transformer, running from top to bottom.

The table on p. 956 gives the properties of some typical magnetic materials; mumetal and ticonal are inventions of the last twenty years, the results of deliberate attempts to develop materials with extreme properties.

Hysteresis Loss; Area of Loop

To calculate the work done in carrying a piece of iron round a hysteresis loop, we adopt the method which we used to calculate the energy stored in the magnetic field of a coil: we consider the back-e.m.f. induced during a change of flux.

Fig 38.8 shows a ring of iron, wound with a uniform magnetizing coil of N turns, and mean length l . If the current through the coil is I , the magnetizing field is

$$H = \frac{NI}{l} \quad (i)$$

And if B is the flux density in the iron, and A its cross-sectional area, the flux through it $= AB$. The flux linkages with the magnetizing coil are therefore

$$\Phi = NAB.$$

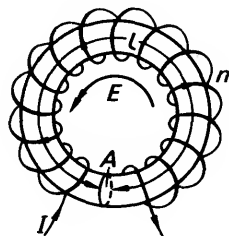


FIG. 38.8. Work done in magnetization of iron.

PROPERTIES OF MAGNETIC MATERIALS

Material	Relative Permeability μ_r	Saturation B_m 10^{-4} T	Remanence B_r 10^{-4} T	Coercivity H_c $A\ m^{-1}$	Hysteresis loss W $J\ m^{-3}$ per cycle	Critical temp. t_c^* $^{\circ}C$	Resistivity† ρ (approx.) $10^{-8}\ \Omega\cdot m$	Applications
SOFT MATERIALS (iron-like, used for electromagnets with varying currents)								
Iron (99.94%)	5500	21500	13000	81.0	500	770	10	Armatures, relays, large transformer cores, telephone diaphragms Small transformer cores, magnetic shields
Nickel . . .	600	6100	3600	272	30	360	7	
Cobalt . . .	240	18000	5000	800	200	1120	10	
Silicon Iron (Stalloy) .	6700	20000	12000	40	350	690	55	
(Fe 96, Si 4) Mumetal (Ni 74, Fe 20, Cu 5, Mn 1)	80000	8500	6000	4	20	—	—	
HARD MATERIALS (steel-like, used for permanent magnets)								
Carbon Steel	—	10000	8000	4800	20000	—	—	Moving-coil instruments, loudspeakers, microphones, telephone ear-piece magnets
Cobalt Steel.	—	—	8000–10000	12000–19200	—	—	—	
Ticonal (Fe 51, Co 24, Ni 14, Al 8, Cu 3)	—	—	12500	44000	—	—	—	

* See page 963.

† The higher the resistivity of a magnetic material, the less the eddy-currents in it when the flux through it is changed; and therefore the less the energy lost as heat.

Now let us suppose that, in a brief time δt , we increase the current by a small amount, and so increase the flux and flux linkages. During the change a back-e.m.f. will be induced in the coil, of magnitude

$$E = \frac{d\Phi}{dt}$$

Hence, from above,

$$E = NA \frac{dB}{dt}$$

To overcome this back-e.m.f., the source of the current I must supply energy to the coil at the rate

$$P = EI$$

Thus the total energy supplied to the coil, in increasing the flux through the iron, is

$$\begin{aligned} \delta W &= P\delta t = EI\delta t \\ &= INA \frac{dB}{dt} \delta t \\ &= INA\delta B \end{aligned}$$

where δB is the increase in flux-density.

Let us now substitute for the current I in terms of the magnetizing field H . From equation (i) we have

$$I = \frac{Hl}{N}$$

Hence

$$\begin{aligned} \delta W &= \frac{lHNA}{N} \delta B \\ &= VH\delta B \end{aligned} \tag{ii}$$

where $V = lA$, is the volume of the iron.

Equation (ii) shows that, in taking the flux from any value B_1 to any other B_2 , the work done is

$$W = V \int_{B_1}^{B_2} HdB \tag{iii}$$

On unit volume of the iron, the work is

$$\int_{B_1}^{B_2} HdB$$

The integral in this expression is the area between the B - H curve and the axis of B . Round the complete hysteresis loop, the work done per unit volume is

$$W = \oint HdB$$

Here the symbol \oint denotes integration round the closed loop (PQRSP in Fig. 38.6); the integral is proportional to the area of the loop.

Subsidiary Hysteresis Loops

When a piece of iron is magnetized, first one way and then the other, it goes round a hysteresis loop even if it is not magnetized to saturation

at any point (Fig. 38.9). The subsidiary loops, *ab* for example, may represent the magnetization of a transformer core by an alternating

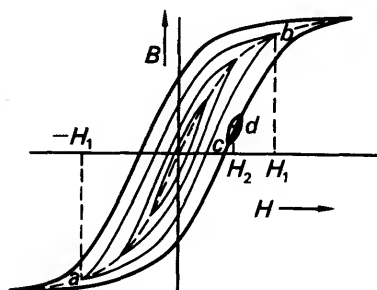


FIG. 38.9. Subsidiary hysteresis loops.

current in the primary winding: the amplitude H_1 of the magnetizing field is proportional to the amplitude of the current. A transformer core is designed so that it is never saturated under working conditions. For, if it were saturated, the flux through it would not follow the changes in primary current; and the e.m.f. induced in the secondary would be less than it should.

The energy dissipated as heat in going round a subsidiary hysteresis loop is proportional to the area of the loop, just as in going round the main one.

Another kind of subsidiary loop is shown at *cd* in the figure. The iron goes round such a loop when the field is varied above and below the value H_2 . This happens in a transformer when the primary carries a fluctuating direct current.

Demagnetization

The only satisfactory way to demagnetize a piece of iron or steel is to carry it round a series of hysteresis loops, shrinking gradually to the origin. If the iron is the core of a toroid, we can do this by connecting the winding to an a.c. supply via a potential divider, as in Fig. 38.10, and reducing the current to zero. Since the iron goes through fifty loops per second, we do not have to reduce the current very slowly.

To demagnetize a loose piece of iron or steel, such as a watch, we merely put it into, and take it out of, a coil of many turns connected to an a.c. supply. As we draw the watch out, it moves into an ever-weakening field, and is demagnetized.

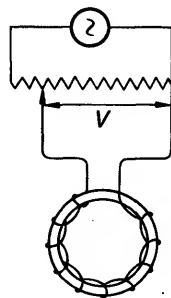


FIG. 38.10. Demagnetization.

Electromagnets and Magnetic Circuits

Consider the electromagnet shown in Fig. 38.11. The magnetic material, which is almost a complete toroid, is wound with N turns

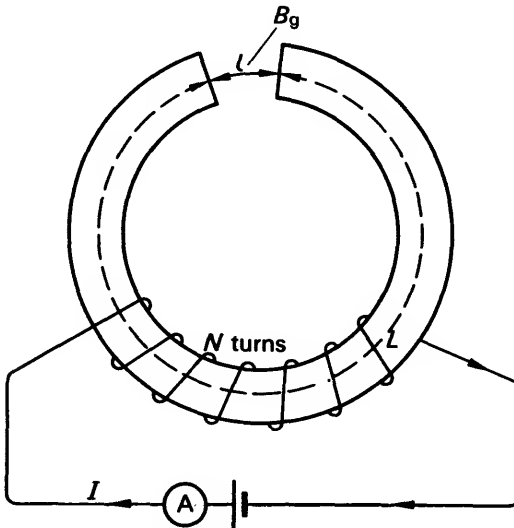


FIG. 38.11. Magnetic circuit.

each carrying a current I . The length of the magnetic material is L and of the small air gap l . We now calculate the induction B_g in the gap between the faces of magnetic material. From $B = \mu H$,

$$B_g = \mu_0 H_g \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and

$$B_m = \mu H_m, \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where H_g is the value of the magnetizing force in the gap, and H_m and B_m are respectively the magnetizing force and flux density inside the material. μ is the permeability of the material.

Taking the closed loop indicated in Fig. 38.11, by Ampere's theorem,

$$H_g l + H_m L = NI.$$

Thus from (1) and (2):

$$\frac{B_g l}{\mu_0} + \frac{B_m L}{\mu} = NI.$$

The total flux Φ round the magnetic circuit is constant. Since the area of cross-section of the air gap is the same as that of the material, it follows that

$$B_g = B_m.$$

$$\therefore B_g \left(\frac{l}{\mu_0} + \frac{L}{\mu} \right) = NI$$

or

$$B_g = \frac{NI}{(l/\mu_0 + L/\mu)} \quad (3)$$

$$\therefore \Phi = B_g A = \frac{NI}{(l/\mu_0 A + L/\mu A)}$$

In general, the value of μ depends on the current flowing in the electro-magnet coil.

EXAMPLE

A toroid made from an iron bar of length 6 cm and of cross-sectional area 4 cm², has an air gap of length 1 cm. If it is wound with 500 turns of wire carrying a current of 20 A, find the flux density in the gap. The material may be assumed to have a permeability of 3000 times that of free space under these conditions (that is $\mu_r = 3000$). If there were no air gap, what would be the induction or flux density in the iron?

We have $N = 500$, $I = 20$ A, $l = 1$ cm = 0.01 m, $\mu = \mu_r \mu_0 = 3000 \times 4\pi \times 10^{-7}$. From (3) above,

$$\begin{aligned} \therefore B_g &= \frac{500 \times 20}{1/\mu_0(l + L/3000)} = \frac{500 \times 20 \times 4\pi \times 10^{-7}}{(0.01 + 0.0002)} \\ &= 1.3 \text{ Wb m}^{-2} \text{ (approx.)} \end{aligned}$$

If $l = 0$, so that there is no air gap, and assuming the value of μ does not alter, then the flux density in the iron is now given by

$$\begin{aligned} B &= \frac{NI}{L/\mu} = \frac{500 \times 20 \times 3000 \times 4\pi \times 10^{-7}}{0.06} \\ &= 53 \text{ Wb m}^{-2}. \end{aligned}$$

Note that even a small air gap considerably reduces the value of the flux density obtained.

Magnetomotive Force. Reluctance.

In equation (3), NI is called the *magnetomotive force*, M.M.F., in the magnetic circuit comprising the iron and air gap. It is analogous to the e.m.f. in an electric circuit. The flux Φ through a cross-section of the circuit, which is given by BA , is analogous to the current.

The quantity $(l/\mu_0 + L/\mu)$ in the denominator in (3) is called the *reluctance* of the magnetic circuit. It is analogous to electrical resistance R , since $R = \rho l/A$ (p. 788). Each term in the expression for the reluctance is due to a separate part of the magnetic circuit. Thus $l/\mu_0 A$ is the reluctance of the air gap and $L/\mu A$ is that of the iron. In the above example, the reluctance of the air gap was $0.01/(4\pi \times 10^{-7} \times A)$ whilst that of the iron was $0.06/(3000 \times 4\pi \times 10^{-7} \times A)$. Thus the reluctance of the air gap is fifty times greater than that of the iron. This accounts for the considerable reduction in the value of B when the air gap is introduced. It should be noted that the reluctances are in series in Fig. 38.11. They are added together, as are series resistances in electric circuits.

Permanent Magnets

A permanent magnet is shown in Fig. 38.12 (i). We apply Ampère's theorem to the circuit shown by the dotted line. Here there are no external currents

flowing so that, using the same notation as in the last section,

$$H_m L + H_g L = 0$$

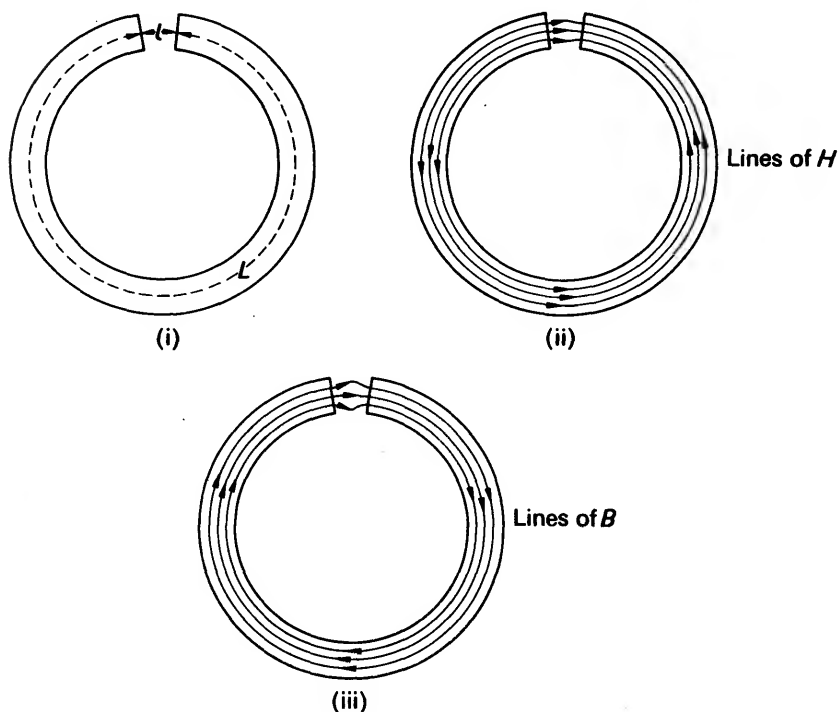


FIG. 38.12. Permanent magnet. B and H .

Also,

$$B_g = \mu H_g \quad \text{and} \quad B_g = B_m,$$

for the same reasons as in the case of electromagnets.

Combining these equations,

$$\therefore \frac{B_m}{H_m} = -\mu_0 \frac{L}{l} \quad (1)$$

This result expresses the fact that B and H are oppositely directed inside the medium. The lines of B and H are shown in Fig. 38.12 (ii) and (iii) respectively. Since $B = \mu H$, the value of μ in the case of this permanent magnet is $-\mu_0 L/l$. Fig. 38.13 shows the points, A or C, representing the magnetic state of the material of a permanent magnet.

The relationship between B and H is provided by the hysteresis loop. Also, B and H inside the material satisfy equation (1), which is represented by the line AOC. The only two points which satisfy both these conditions are A and C, the points of intersection of the line with the curve. At A or C, H is oppositely directed to B and hence μ is negative. For any given shape of toroidal magnet, one can predict, using (1), the greatest possible induction in the air gap.

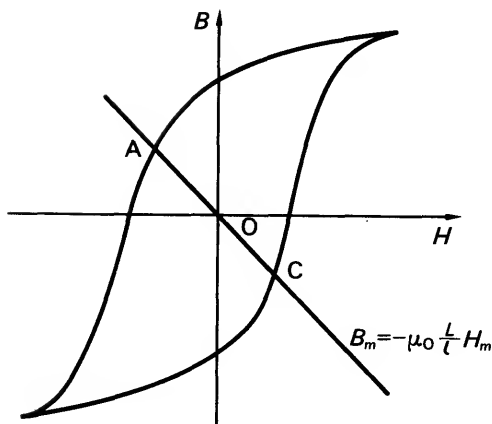


FIG. 38.13. Permanent magnet.

Diamagnetism, Paramagnetism, Ferromagnetism

We have already mentioned that the magnetic properties of materials are due to circulating and spinning electrons within the atoms.

Diamagnetism

If a magnetic field is produced in the neighbourhood of a magnetic material, a changing flux occurs in the current loops within the atoms. An e.m.f. or electric field will then be set up which causes the electrons to alter their motions, so that an extra or induced current is produced. By Lenz's law, this current gives rise to a magnetic field which *opposes* the applied magnetic field H . Thus the induced magnetization will be in the opposite direction to H , that is, M/H is negative. Hence the susceptibility χ is negative. This phenomenon is called *diamagnetism*. For a diamagnetic material, χ is generally very small, about -0.000015 for bismuth, for example. The relative permeability, μ , which is given by $\mu_r = 1 + \chi$, is thus generally slightly less than 1. All substances have a diamagnetic contribution to their susceptibility, since the induced currents always oppose the applied field. In many substances, the diamagnetism is completely masked by another magnetic phenomenon (p. 953).

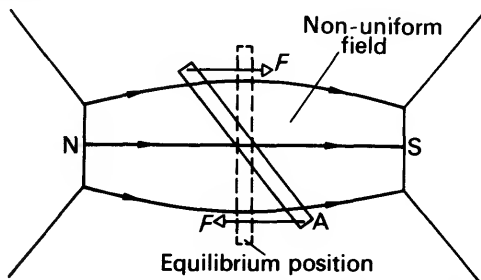


FIG. 38.14. Rod of diamagnetic material in strong field.

If a rod of diamagnetic material is placed in a non-uniform magnetic field, it will settle at right angles to the field. Fig. 38.14 shows the specimen slightly

displaced from this position. The magnetization will oppose the applied field so that the end A will now effectively be a weak S pole. It will then experience a force as shown by the arrow, so that a restoring couple turns the specimen back to its position at right angles to the field.

It should be noted that diamagnetism is a natural 'reaction' to an applied magnetic field and that it is independent of temperature.

Paramagnetism

In contrast to bismuth, a rod of a material such as platinum will settle along the same direction as the applied magnetic field. Further, the induced magnetism will be in the same direction as the field. Fig. 38.15. Platinum is an example of a *paramagnetic* material. The susceptibility, χ , of a paramagnetic substance is very small and positive, +0.0001 for example, so that its relative permeability μ_r is very slightly greater than 1 from $\mu_r = 1 + \chi$.

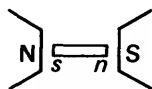


FIG. 38.15. Rod of paramagnetic material in strong field.

As we have already mentioned, atoms contain circulating and spinning electrons. Each electron possesses a resultant magnetic moment on account of its orbital motion and its spin motion. In a diamagnetic atom, all these contributions to the magnetic moment cancel. In a paramagnetic atom, however, there is a resultant magnetic moment. Generally, the thermal motions of the atoms will cause these magnetic moments to be oriented purely at random and there will be no resultant magnetization. If, however, a field is applied, each atomic moment will try to set in the direction of the field but the thermal motions will prevent complete alignment. In this case there will be overall weak magnetization in the direction of the applied field. This accounts for the phenomenon of paramagnetization.

It is clear that paramagnetism is temperature dependent. At low temperatures, the thermal motions will be less successful at preventing the alignment of the atomic moments and so the susceptibility will be larger. At higher temperatures thermal motion will make alignment difficult. At very high temperatures, the material may become diamagnetic, for the diamagnetic contribution to χ is not affected by temperature whilst the paramagnetic contribution falls.

Ferromagnetism. Magnetic Domain Theory

A ferromagnetic material has a very high value of susceptibility, χ , and hence of relative permeability, μ_r . The value of μ_r can be several thousands. Like a paramagnetic material, the magnetization is in the direction of the applied field and a rod of a ferromagnetic material will align itself along the field.

In a paramagnetic substance which is not subjected to a magnetic field, the magnetic moments are oriented purely at random due to the thermal vibrations. In a ferromagnetic material, however, strong 'interactions' are present between the moments, the nature of which requires quantum theory to understand it and is outside the scope of this work. These cause neighbouring moments to align, even in the absence of an applied field, with the result that tiny regions of very strong magnetism are obtained inside the unmagnetized material called *magnetic domains*. Above a critical temperature called the *Curie point*, ferromagnetics become paramagnetics (see Table, p. 956).

Domain Formation

A crystal of ferromagnetic material is shown in Fig. 38.16 (i). If all the domains were aligned completely, the material would behave like one enormous domain and the energy in the magnetic field outside is then considerable, as represented by the flux shown. Now all physical systems settle in equilibrium when their energy is a minimum. Fig. 38.16(ii) is therefore more stable than Fig. 38.16(i)

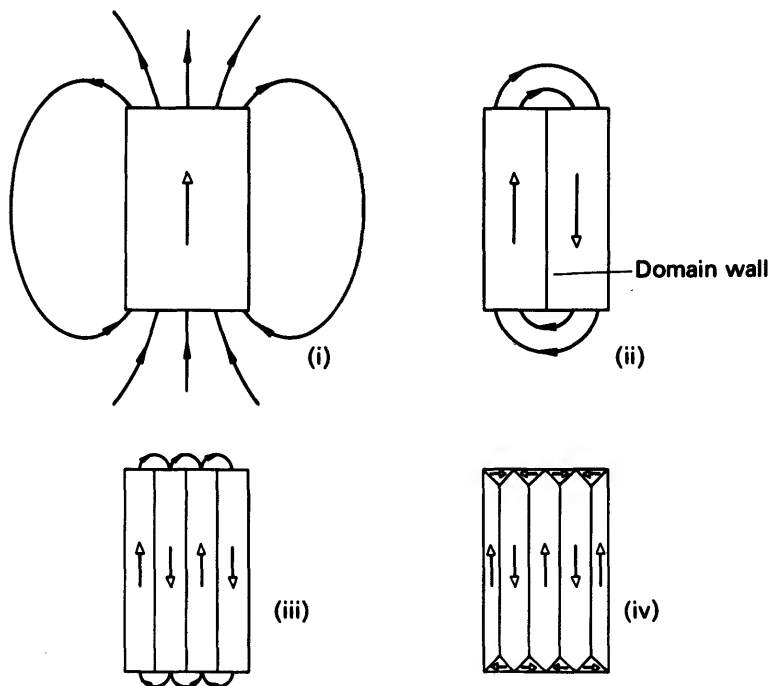


FIG. 38.16. Magnetic domains.

because the external magnetic field energy is less. Thus the domains grow in the material, as shown in Fig. 38.16 (iii) and (iv). The region between two domains, where the magnetization changes direction, is called a *domain wall* and also contains energy. When the formation of a new domain wall requires more energy than is gained by the reduction in the external magnetic field, no more domains are formed. Thus there is a limit to the number of domains formed. This occurs when the volume of the domains is of the order 10^{-4} cm^3 or less.

Domains and Magnetization

Some of the phenomena in magnetization of ferromagnetic material can now be explained. In an unmagnetized specimen, the domains are oriented in different directions. The net magnetization is then zero. If a small magnetic field H is applied, there is some small rotation of the magnetization within the domains, which produces an overall component of magnetization in the direction of H . This occurs in the region AB of the magnetization-field (M - H) curve shown in Fig. 38.17.

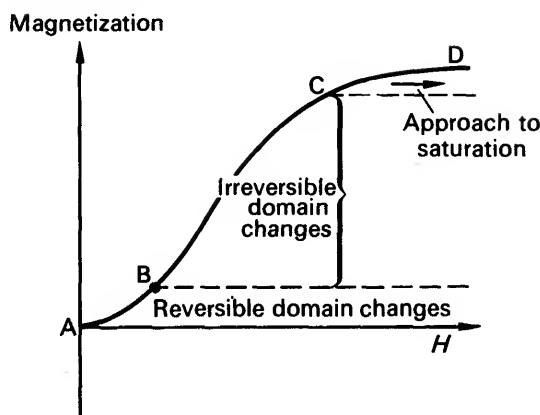


FIG. 38.17. Domain movement in magnetization.

If the field H is removed, the domain magnetization returns to its original direction. Thus the magnetization returns to zero. The changes in the part AB of the curve are hence reversible. If the field H is increased beyond B in the region BC, the magnetization becomes greater. On removal of the field the magnetization does not return to zero, and so remanence occurs. Along BC, then, irreversible changes take place; the domains grow in the direction of the field, by movement of domain walls, at the expense of those whose magnetization is in the opposite direction. At very high applied fields H there is complete alignment of the domains and so the magnetization M approaches 'saturation' along CD.

EXERCISES 38

1. Define *intensity of magnetization*, *susceptibility* and *permeability*. Two substances, A and B, have relative permeabilities slightly greater and slightly less than unity respectively. What does this signify about their magnetic properties? To what group of magnetic substances do A and B each belong?

A soft iron ring of cross-sectional diameter 8 cm and mean circumference 200 cm has 400 turns of wire wound uniformly on it. Calculate the current necessary to produce magnetic flux to the value of 5×10^{-4} Wb if the relative permeability of the iron in the condition stated is 1800. Why is it not possible to say from this information what the flux would be if the current were reduced to 1/10 of its calculated value? (L.)

2. Define *permeability (relative permeability)* of a magnetic substance.

Indicate the orders of magnitude of the relative permeabilities of ferromagnetic, paramagnetic and diamagnetic substances respectively. Discuss in relation to its (relative) permeability one aspect of the behaviour of a specimen of each of these substances when placed in turn in the same magnetic field.

An iron ring of mean circumference 30 cm and of area of cross-section 1.5 cm^2 has 240 turns of wire uniformly wound on it, through which passes a current of 2 A. If the flux in the iron is found to be 7.5×10^{-4} Wb, find the relative permeability of the iron. (L.)

3. Give diagrams to illustrate the distribution of the lines of induction when a sphere of (a) soft iron, (b) bismuth (which is diamagnetic) is placed in a magnetic

field, initially uniform. Compare and contrast the magnetic properties of these materials.

The magnetic induction (flux density) in a uniformly magnetized specimen of cast iron is 0.3 Wb m^{-2} when the strength of the magnetizing field is 1000 A m^{-1} . Find (i) the intensity of magnetization, (ii) the relative permeability, (iii) the magnetic susceptibility of the specimen. (L.)

4. What is meant by *intensity of magnetization* and *hysteresis*? Describe an experiment from the observations of which a graph of intensity of magnetization (M) against magnetizing field strength (H) over a complete hysteresis cycle may be plotted for a ferromagnetic material. What information may be obtained from the graph by measuring (a) the area enclosed by the loop, (b) the intercept on the M axis, (c) the intercept on the H axis, (d) the slope of a line joining a point on the graph to the origin? (L.)

5. Define *intensity of magnetization* M and *magnetic susceptibility* χ_m , and with the aid of a diagram explain what is meant by *hysteresis*.

Sketch a graph showing how M varies with the applied magnetizing field as the field applied to a specimen of soft iron, initially unmagnetized, is slowly increased from zero. On the same diagram sketch the corresponding graph for steel.

By reference to your earlier account state, with reasons, desirable magnetic properties of materials to be used as (a) the core of an electromagnet, (b) the core of a transformer, (c) a permanent magnet. (L.)

6. Define *magnetic moment* and explain why this concept is useful in the study of magnetism.

A bar of magnetic material is magnetized in the uniform field inside a solenoid. How would you study experimentally the changes in its magnetic moment when the magnetizing current is varied in magnitude and direction?

How would you represent the results of your measurements graphically? Give freehand sketches of the graphs you would expect to obtain for (i) material suitable for the core of a transformer, (ii) material suitable for a permanent magnet. Point out in each case the features of the graph which indicate that the material is suitable for its purpose. (O. & C.)

7. Define *susceptibility*, *permeability*, and state the relation between them.

Describe briefly how you would test if a small rod is diamagnetic, paramagnetic, or ferromagnetic. Draw intensity of magnetization graphs for each type of material and comment on their special features. (N.)

8. Explain what is meant by a *ferromagnetic material*. Give a general account of the wide range of magnetic properties exhibited by different ferromagnetic materials and indicate the practical applications of these properties. (N.)

chapter thirty-nine

A.C. Circuits. Electromagnetic Waves

Measurement of A.C.

If an alternating current (a.c.) is passed through a moving-coil meter, the pointer does not move. The coil is urged clockwise and anticlockwise at the frequency of the current—50 times per second if it is drawn from the British grid—and does not move at all. In a delicate instrument the pointer may be seen to vibrate, with a small amplitude. Instruments for measuring alternating currents must be so made that the pointer deflects the same way when the current flows through the instrument in either direction.

Moving-iron Instrument

A fairly common type of such instrument, called the moving-iron type, is shown in Fig. 39.1. It consists of two iron rods XY, PQ, surrounded by a coil which carries the current. The coil is fixed to the framework of the meter, and so is one of the rods PQ. The other rod is attached to an axle, which also carries the pointer, its motion is controlled by hair-springs. When a current flows through the coil, it

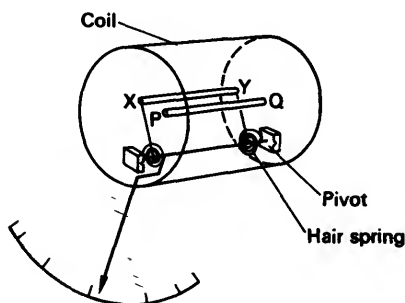


FIG. 39.1. Moving-iron meter, repulsion type.

magnetizes the rods, so that their adjacent ends have the same polarity. The polarity of each pair of ends reverses with the current, but whichever direction the current has, the iron rods repel each other. The force on the pivoted rod is therefore always in the same direction, and the pointer is deflected through an angle which is proportional to the average force. To a fair approximation, the magnetization of the rods at any instant is proportional to the current at that instant; the force

between the rods is therefore roughly proportional to the square of the current. The deflection of the pointer is therefore roughly proportional to the average value of the square of the current.

Hot-wire Instrument

Another type of 'square law' instrument is the hot-wire ammeter (Fig. 39.2). In it the current flows through a fine resistance-wire XY, which it heats. The wire warms up to such a temperature that it loses heat—mainly by convection—at a rate equal to the average rate at which heat is developed in the wire. The rise in temperature of the wire makes it expand and sag; the sag is taken up by a second fine wire PQ,

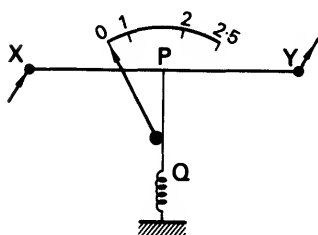


FIG. 39.2. Hot wire meter.

which is held taut by a spring. The wire PQ passes round a pulley attached to the pointer of the instrument, which rotates as the wire XY sags. The deflection of the pointer is roughly proportional to the average rate at which heat is developed in the wire XY; it is therefore roughly proportional to the average value of the square of the alternating current, and the scale is a square-law one.

Root-mean-square value of A.C.

On p. 907 we saw that an alternating current I varied sinusoidally; that is, it could be represented by the equation $I = I_m \sin \omega t$, where I_m was the peak (maximum) value of the current. In commercial practice, alternating currents are always measured and expressed in terms of their *root-mean-square (r.m.s.)* value, which we shall now consider.

Consider two resistors of equal resistance R , one carrying an alternating current and the other a direct current. Suppose both are dissipating the same power, P , as heat. The root-mean-square (r.m.s.) value of the alternating current, I_r , is defined as equal to the direct current, I_d . Thus:

the root-mean-square value of an alternating current is defined as that value of steady current which would dissipate heat at the same rate in a given resistance.

Since the power dissipated by the direct current is

$$P = I_d^2 R,$$

our definition means that, in the a.c. circuit,

$$P = I_r^2 R. \quad (1)$$

Whatever the wave-form of the alternating current, if I is its value at any instant, the power which it delivers to the resistance R at that instant is $I^2 R$. Consequently, the average power P is given by

$$\begin{aligned} P &= \text{average value of } (I^2 R) \\ &= R \times \text{average value of } (I^2), \end{aligned}$$

since R is a constant. Therefore, by equation (1),

$$I_r^2 R = R \times \text{average value of } (I^2)$$

or
$$I_r^2 = \text{average value of } (I^2). \tag{2}$$

The average value of (I^2) is called the mean-square current ; the meaning of the term is illustrated for a non-sinusoidal current in Fig. 39.3 (i).

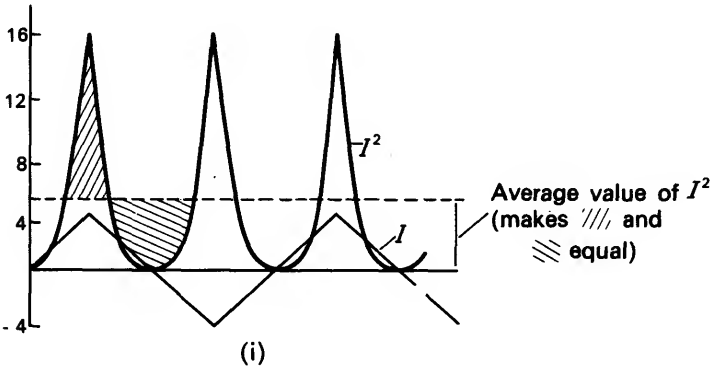


FIG. 39.3 (i). Mean-square values.

$$\therefore I_r = \sqrt{\text{average value of } (I^2)}. \tag{3}$$

For a sinusoidal current, the average value of I^2 is $I_m^2/2$, where I_m is the maximum value of the current

$$\therefore I_r = \frac{I_m}{\sqrt{2}}. \tag{4}$$

Equation (3) shows the origin of the term ‘root-mean-square’. We therefore require a meter whose deflection measures not the current through it but the average value of the square of the current. As we have already seen, moving-iron and hot-wire meters have just this property (p. 968).

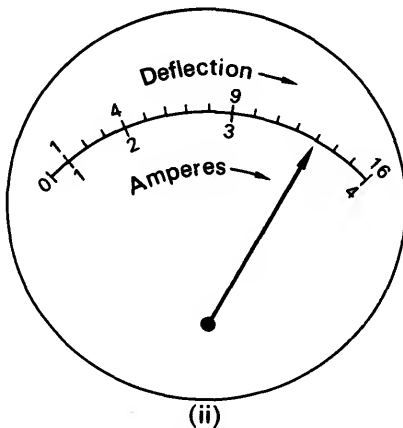


FIG. 39.3 (ii). Scale of A.C. ammeter.

For convenience, such meters are scaled to read amperes, not (amperes)², as in Fig. 39.3 (ii). The scale reading is then proportional to the square-root of the deflection, and indicates directly the root-mean-square value of the current I_r . An a.c. meter of the moving-iron or hot-wire type can be calibrated in direct current, as shown in Fig. 39.3 (ii). This follows at once from the definition of the r.m.s. value of current.

A.C. through a Capacitor

In many radio circuits, resistors, capacitors, and coils are present. An alternating current can flow through a resistor, but it is not obvious at first that it can flow through a capacitor. This can be demonstrated, however, by wiring a capacitor, of capacitance one or more microfarads, in series with a neon lamp, and connecting them to the a.c. mains via a plug. In place of the neon lamp we could use a mains filament lamp of low rating, such as 25 watts. The lamp lights up, showing that a current is flowing through it.

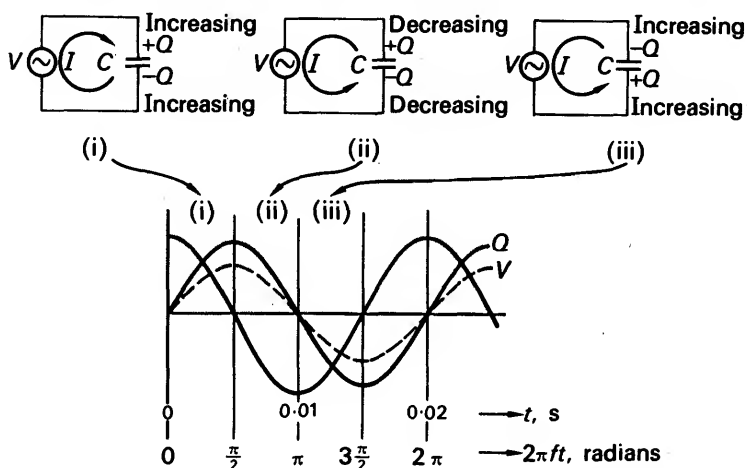


FIG. 39.4. Flow of A.C. through capacitor, frequency 50 Hz.

The current flows because the capacitor plates are being continually charged, discharged, and charged the other way round by the alternating voltage of the mains (Fig. 39.4). The current thus flows round the circuit, and can be measured by an a.c. milliammeter inserted in any of the connecting wires.

To find the amplitude of the current, let us denote the amplitude of the voltage by V_m , and its frequency by f . Then, as in equation (3), p. 907, the instantaneous voltage at any time t is

$$V = V_m \sin 2\pi ft.$$

If C is the capacitance of the capacitor, then the charge Q on its plates is

$$Q = CV,$$

whence

$$Q = CV_m \sin 2\pi ft.$$

The current flowing at any instant, I , is equal to the rate at which charge is accumulating on the capacitor plates. Thus

$$\begin{aligned} I &= \frac{dQ}{dt} \\ &= \frac{d}{dt}(CV_m \sin 2\pi ft) \\ &= 2\pi f CV_m \cos 2\pi ft. \end{aligned} \quad (1)$$

Equation (1) shows that the amplitude of the current is $2\pi f CV_m$; proportional to the frequency, the capacitance, and the voltage amplitude. These results are easy to explain. The greater the voltage, or the capacitance, the greater the charge on the plates, and therefore the greater the current required to charge or discharge the capacitor. And the higher the frequency, the more rapidly is the capacitor charged and discharged, and therefore again the greater the current.

A more puzzling feature of equation (1) is the factor giving the time variation of the current, $\cos 2\pi ft$. It shows that the current varies a quarter-cycle out of step with the voltage—or, as is more often said, $\pi/2$ out of phase. Fig. 39.4 shows this variation, and also helps to explain it. When the voltage is a maximum, so is the charge on the capacitor. It is therefore not charging and the current is zero. When the voltage starts to fall, the capacitor starts to discharge; the rate of discharging, or current, reaches its maximum when the capacitor is completely discharged and the voltage across it is zero.

A.C. through an Inductor

Since a coil is made from conducting wire, we have no difficulty in seeing that an alternating current can flow through it. However, if the coil has appreciable self-inductance, the current is less than would flow through a non-inductive coil of the same resistance. We have already seen how self-inductance opposes changes of current; it must therefore oppose an alternating current, which is continuously changing.

Let us suppose that the resistance of the coil is negligible, a condition which can easily be satisfied in practice. We can simplify the theory by considering first the current, and working back to find the potential difference across the coil. Let us therefore denote the current by

$$I = I_m \sin 2\pi ft, \quad (1)$$

where I_m is its amplitude (Fig. 39.5). If L is the inductance of the coil, the changing current sets up a back-e.m.f. in the coil, of magnitude

$$E = L \frac{dI}{dt}.$$

Thus

$$\begin{aligned} E &= L \frac{d}{dt}(I_m \sin 2\pi ft) \\ &= 2\pi f L I_m \cos 2\pi ft. \end{aligned}$$

To maintain the current, the applied supply voltage must be equal to the back-e.m.f. The voltage applied to the coil must therefore be given by

$$V = 2\pi f L I_m \cos 2\pi f t. \quad (2)$$

Equation (2) shows that the amplitude of the voltage across a pure inductance (without resistance) is proportional to the frequency, the magnitude of the self-inductance, and the amplitude of the current. The reader is left to work out the physical explanation of these three facts for himself.

The voltage across a pure inductance is a quarter-cycle, or $\pi/2$ radians, out of phase with the current. Fig. 39.5 shows this relationship, which follows from comparing equations (1) and (2). It can be explained by considering the relationship between the back-e.m.f., and the changing current, which the reader should do for himself.

The phase relationship between current and voltage for a pure inductance is, however, different from that for a capacitor. For, as Fig. 39.5 shows, the current through an inductance lags $\pi/2$ behind the voltage; that is to say, it passes its maxima a quarter-cycle later than the voltage. On the other hand, in a capacitor the current passes its maxima a quarter-cycle ahead of the voltage (Fig. 39.4): we say that it leads the voltage by $\pi/2$.

Reactance

The term reactance is used to denote the opposition which an inductance or capacitor offers to the passage of an alternating current. We do not use the term resistance for this, because we have already defined resistance as the property of a circuit which enables it to dissipate electrical power as heat; and we have just seen that a capacitor or a pure inductance dissipates no electrical power at all.

We define the reactance of an inductor L , or capacitor C , by the equation

$$\text{reactance} = \frac{\text{amplitude of voltage across } L \text{ or } C}{\text{amplitude of current through it}}$$

Denoting reactance by X , we have

$$X = \frac{V_m}{I_m}.$$

Since V_m is measured in volts, and I_m in amperes, the reactance X is expressed in ohms.

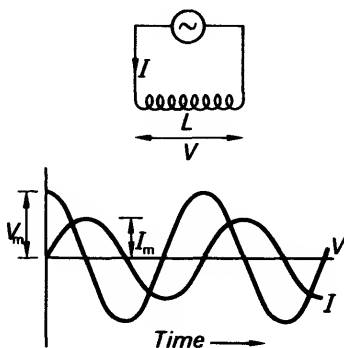


FIG. 39.5. Passage of a.c. through a coil.

We have already seen (p. 971) that the amplitude of the current through a capacitor is given by

$$I_m = 2\pi f C V_m.$$

The reactance of the capacitor is therefore

$$X_C = \frac{V_m}{I_m} = \frac{1}{2\pi f C}.$$

X_C is in ohms when f is in Hz (cycles per second), and C in farads.

The amplitude of the voltage across a pure inductor is

$$V_m = 2\pi f L I_m.$$

Hence the reactance of the inductor is

$$X_L = \frac{V_m}{I_m} = 2\pi f L.$$

It is in ohms when f is in Hz, and L is in henrys (H).

For convenience we often write $\omega = 2\pi f$.

The quantity ω is called the angular frequency of the current and voltage. It is expressed in radians per second. Then an alternating voltage, for example, may be written as

$$V = V_m \sin \omega t.$$

And reactances become

$$X_L = \omega L \quad X_C = \frac{1}{\omega C}.$$

EXAMPLE

1. A capacitor C of $1 \mu\text{F}$ is used in a radio circuit where the frequency is 1000 Hz and the current flowing is 2 mA. Calculate the voltage across C .

$$\text{Reactance, } X_C = \frac{1}{2\pi f C} = \frac{1}{2\pi \times 1000 \times 1/10^6} = 159 \text{ ohms (approx.)}$$

$$\therefore V = I X_C = \frac{2}{1000} \times 159 = 0.32 \text{ V (approx.)}$$

2. An inductor of 2 H and negligible resistance is connected to a 12 V mains supply, $f = 50$ Hz. Find the current flowing.

$$\text{Reactance, } X_L = 2\pi f L = 2\pi \times 50 \times 2 = 628 \text{ ohms.}$$

$$\therefore I = \frac{V}{X_L} = \frac{12}{628} \text{ A} = 19 \text{ mA (approx.)}$$

Vector Diagram

In the Mechanics section of this book, it is shown that a quantity which varies sinusoidally with time may be represented as the projection of a rotating vector (p. 44). Alternating currents and voltages may therefore be represented in this way. Fig. 39.6 shows, on the left, the vectors representing the current through a capacitor, and

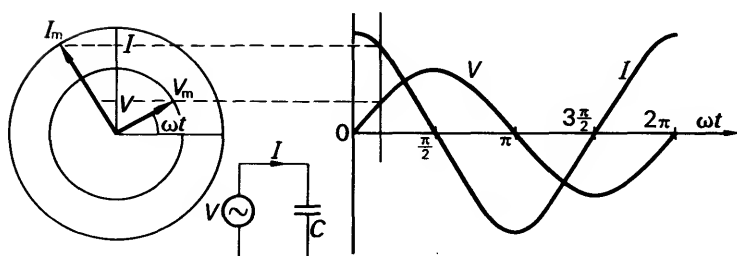


FIG. 39.6. Vector diagram for capacitor.

the voltage across it. Since the current leads the voltage by $\pi/2$, the current vector I is displaced by 90° ahead of the voltage vector V .

Fig. 39.7 shows the vector diagram for a pure inductor. In drawing it, the voltage has been taken as $V = V_m \sin \omega t$, and the current drawn

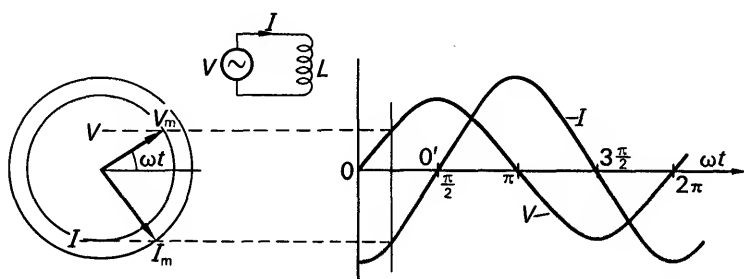


FIG. 39.7. Vector diagram for pure inductance.

lagging $\pi/2$ behind it. This enables the diagram to be readily compared with that for a capacitor. To show that it is essentially the same as Fig. 39.5, we have only to shift the origin by $\pi/2$ to the right, from 0 to $0'$.

SERIES CIRCUITS

L and R in series

Consider an inductor L in series with resistance R , with an alternating voltage V (r.m.s.) of frequency f connected across both components (Fig. 39.8 (i)).

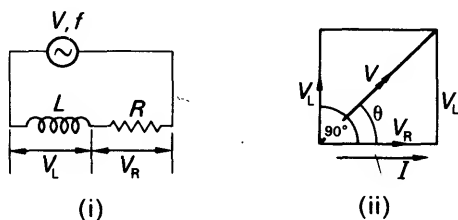


FIG. 39.8. Inductance and resistance in series.

The sum of the respective voltages V_L and V_R across L and R is equal to V . But the voltage V_L leads by 90° on the current I , and the voltage V_R is in phase with I (see page 972). Thus the two voltages can be drawn to scale as shown in Fig. 39.8 (ii), and hence, by Pythagoras' theorem, it follows that the vector sum V is given by

$$V^2 = V_L^2 + V_R^2.$$

But $V_L = IX_L$, $V_R = IR$.

$$\therefore V^2 = I^2 X_L^2 + I^2 R^2 = I^2 (X_L^2 + R^2),$$

$$\therefore I = \frac{V}{\sqrt{X_L^2 + R^2}}. \quad (i)$$

Also, from Fig. 39.8 (ii), the current I lags on the applied voltage V by an angle θ given by

$$\tan \theta = \frac{V_L}{V_R} = \frac{IX_L}{IR} = \frac{X_L}{R}. \quad (ii)$$

From (i), it follows that the 'opposition' Z to the flow of alternating current is given in ohms by

$$Z = \frac{V}{I} = \sqrt{X_L^2 + R^2}. \quad (iii)$$

This 'opposition', Z , is known as the *impedance* of the circuit.

EXAMPLE

An iron-cored coil of 2 H and 50 ohms resistance is placed in series with a resistor of 450 ohms, and a 100 V, 50 Hz, a.c. supply is connected across the arrangement. Find the current flowing in the coil.

The reactance $X_L = 2\pi fL = 2\pi \times 50 \times 2 = 628$ ohms.

Total resistance $R = 50 + 450 = 500$ ohms.

$$\therefore \text{circuit impedance } Z = \sqrt{X_L^2 + R^2} = \sqrt{628^2 + 500^2} = 803 \text{ ohms}$$

$$\therefore I = \frac{V}{Z} = \frac{100}{803} \text{ A} = 12.5 \text{ mA (approx.)}$$

C and R in series

A similar analysis enables the impedance to be found of a capacitance C and resistance R in series. Fig. 39.9 (i). In this case the voltage V_C

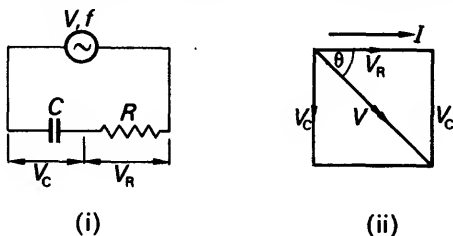


FIG. 39.9. Capacitance and resistance in series.

across the capacitor lags by 90° on the current I (see p. 971), and the voltage V_R across the resistance is in phase with the current I . As the vector sum is V , the applied voltage, it follows by Pythagoras' theorem that

$$V^2 = V_C^2 + V_R^2 = I^2 X_C^2 + I^2 R^2 = I^2 (X_C^2 + R^2),$$

$$\therefore I = \frac{V}{\sqrt{X_C^2 + R^2}}. \quad (i)$$

Also, from Fig. 39.9 (ii), the current I leads on V by an angle θ given by

$$\tan \theta = \frac{V_C}{V_R} = \frac{IX_C}{IR} = \frac{X_C}{R}. \quad (ii)$$

It follows from (i) that the impedance Z of the C - R series circuit is

$$Z = \frac{V}{I} = \sqrt{X_C^2 + R^2}.$$

It should be noted that although the impedance formula for a C - R series circuit is of the same mathematical form as that for a L - R series circuit, the current in the former case leads on the applied voltage but the current in the latter case lags on the applied voltage.

L , C , R in series

The most general series circuit is the case of L , C , R in series (Fig. 39.10 (i)). The vector diagram has V_L leading by 90° on V_R , V_C lagging by 90° on V_R , with the current I in phase with V_R (Fig. 39.10 (ii)). If V_L is greater than V_C , their resultant is $(V_L - V_C)$ in the direction of V_L ,

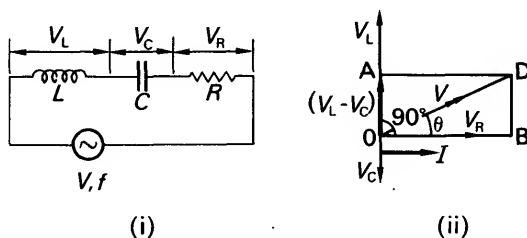


FIG. 39.10. L , C , R in series.

as shown. Thus, from Pythagoras' theorem for triangle ODB , the applied voltage V is given by

$$V^2 = (V_L - V_C)^2 + V_R^2.$$

But $V_L = IX_L$, $V_C = IX_C$, $V_R = IR$.

$$\therefore V^2 = (IX_L - IX_C)^2 + I^2 R^2 = I^2 [(X_L - X_C)^2 + R^2],$$

$$\therefore I = \frac{V}{\sqrt{(X_L - X_C)^2 + R^2}}. \quad (i)$$

Also, I lags on V by an angle θ given by

$$\tan \theta = \frac{DB}{OB} = \frac{V_L - V_C}{V_R} = \frac{IX_L - IX_C}{IR} = \frac{X_L - X_C}{R}. \quad (ii)$$

Resonance in the L, C, R series circuit

From (i), it follows that the impedance Z of the circuit is given by

$$Z = \sqrt{(X_L - X_C)^2 + R^2}.$$

The impedance varies as the frequency, f , of the applied voltage varies, because X_L and X_C both vary with frequency. Since $X_L = 2\pi fL$, then $X_L \propto f$, and thus the variation of X_L with frequency is a straight line passing through the origin (Fig. 39.11 (i)). Also, since $X_C = 1/2\pi fC$, then $X_C \propto 1/f$, and thus the variation of X_C with frequency is a curve approaching the two axes (Fig. 39.11 (i)). The resistance R is independent of frequency, and is thus represented by a line

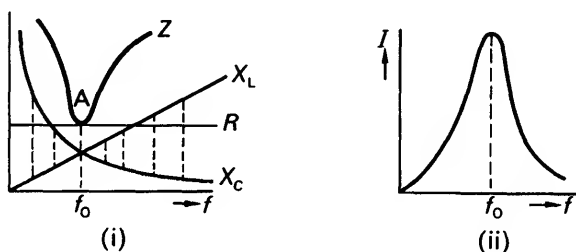


FIG. 39.11. Resonance curves.

parallel to the frequency axis. The difference $(X_L - X_C)$ is represented by the dotted lines shown in Fig. 39.11 (i), and it can be seen that $(X_L - X_C)$ decreases to zero for a particular frequency f_0 , and thereafter increases again. Thus, from $Z = \sqrt{(X_L - X_C)^2 + R^2}$, the impedance diminishes and then increases as the frequency f is varied. The variation of Z with f is shown in Fig. 39.11 (i), and since the current $I = V/Z$, the current varies as shown in Fig. 39.11 (ii). Thus the current has a maximum value at the frequency f_0 , and this is known as the *resonant frequency* of the circuit.

The magnitude of f_0 is given by $X_L - X_C = 0$, or $X_L = X_C$.

$$\therefore 2\pi f_0 L = \frac{1}{2\pi f_0 C} \quad \text{or} \quad 4\pi^2 LC f_0 = 1.$$

$$\therefore f_0 = \frac{1}{2\pi\sqrt{LC}}$$

At frequencies above and below the resonant frequency, the current is less than the maximum current, see Fig. 39.11 (ii), and the phenomenon is thus basically the same as the forced and resonant vibrations obtained in Sound or Mechanics (p. 581).

The series resonance circuit is used for tuning a radio receiver. In this case the incoming waves of frequency f say from a distant transmitting station induces a varying voltage in the aerial, which in turn induces a voltage V of the same frequency in a coil and capacitor circuit in the receiver (Fig. 39.12). When the capacitance C is varied the resonant frequency is changed; and at one setting of C the resonant frequency becomes f , the frequency of the incoming waves. The maximum current is then obtained, and the station is now heard very loudly.

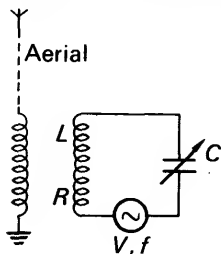


FIG. 39.12.
Tuning a receiver.

Power in A.C. circuits

Resistance R . The power absorbed generally is $P = IV$. In the case of a resistance, $V = IR$, and $P = I^2R$. The variation of power is shown in Fig. 39.13 (i), from which it follows that the average power absorbed $P = I_m^2R/2$, where $I_0 = I_m$ = the peak (maximum) value of the current. Since the r.m.s. value of the current is $I_m/\sqrt{2}$, it follows that

$$P = I^2R,$$

where I is the r.m.s. value (see p. 968).

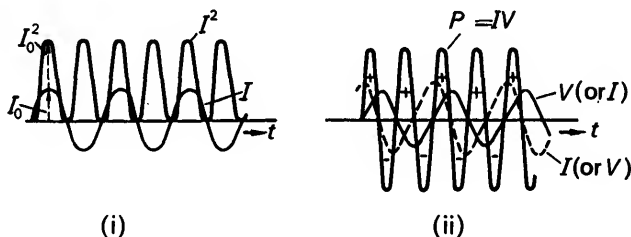


FIG. 39.13. Power in A.C. circuits.

Inductance L . In the case of a pure inductor, the voltage V across it leads by 90° on the current I . Thus if $I = I_m \sin \omega t$, then $V = V_m \sin (90^\circ + \omega t) = V_m \cos \omega t$. Hence, at any instant,

$$\text{power absorbed} = IV = I_m V_m \sin \omega t \cdot \cos \omega t = \frac{1}{2} I_m V_m \sin 2\omega t.$$

The variation of power, P , with time t is shown in Fig. 39.13 (ii); it is a sine curve with an average of zero. Hence no power is absorbed in a pure inductance. This is explained by the fact that on the first quarter of the current cycle, power is absorbed (+) in the magnetic field of the coil (see p. 925). On the next quarter-cycle the power is returned (–) to the generator, and so on.

Capacitance. With a pure capacitance, the voltage V across it lags by 90° in the current I (p. 971). Thus is $I = I_m \sin \omega t$,

$$V = V_m \sin (\omega t - 90^\circ) = -V_m \cos \omega t.$$

Hence, numerically,

$$\text{power at an instant, } P, = IV = I_m V_m \sin \omega t \cos \omega t = \frac{I_m V_m}{2} \sin 2\omega t.$$

Thus, as in the case of the inductance, *the power absorbed in a cycle is zero* (Fig. 39.13 (ii)). This is explained by the fact that on the first quarter of the cycle, energy is stored in the electrostatic field of the capacitor. On the next quarter the capacitor discharges, and the energy is returned to the generator.

Formulae for A.C. Power

It can now be seen that, if I is the r.m.s. value of the current in amps in a circuit containing a resistance R ohms, the power absorbed is $I^2 R$ watts. Care should be taken to exclude the inductances and capacitances in the circuit, as no power is absorbed in them.

If the voltage V across a circuit leads by an angle θ on the current I , the voltage can be resolved into a component $V \cos \theta$ in phase with the current, and a voltage $V \sin \theta$ perpendicular to the current (Fig. 39.14). The former component, $V \cos \theta$, represents that part of the

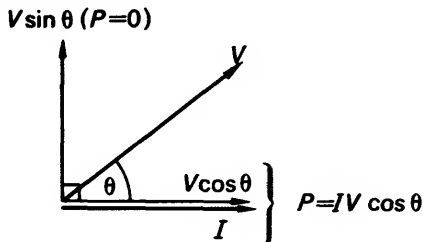


FIG. 39.14. Power absorbed.

voltage across the total *resistance* in the circuit, and hence the power absorbed is

$$P = IV \cos \theta.$$

The component $V \sin \theta$ is that part of the applied voltage across the total inductance and capacitance. Since the power absorbed here is zero, it is sometimes called the 'wattless component' of the voltage.

EXAMPLE

A circuit consists of a capacitor of $2 \mu\text{F}$ and a resistor of 1000 ohms. An alternating e.m.f. of 12 V (r.m.s.) and frequency 50 Hz is applied. Find (1) the current flowing, (2) the voltage across the capacitor, (3) the phase angle between the applied e.m.f. and current, (4) the average power supplied.

The reactance X_C of the capacitor is given by

$$X_C = \frac{1}{2\pi fC} = \frac{1}{2\pi \times 50 \times 2/10^6} = 1590 \text{ ohms (approx.)}$$

\therefore total impedance $Z = \sqrt{R^2 + X_C^2} = \sqrt{1000^2 + 1600^2} = 1880$ ohms (approx.).

(1) \therefore current, $I = \frac{V}{Z} = \frac{12}{1880} = 6.4 \times 10^{-3}$ A.

(2) voltage across, $V_C = IX_C = \frac{12}{1880} \times 1590 = 10.2$ V (approx.).

(3) The phase angle ϕ is given by

$$\tan \phi = \frac{X_C}{R} = \frac{1590}{1000} = 1.59$$

$$\therefore \phi = 58^\circ \text{ (approx.)}$$

(4) Power supplied $= I^2 R = \left(\frac{12}{1880} \right)^2 \times 1000 = 0.04$ W (approx.).

ELECTROMAGNETIC WAVES

Alternating current circuits containing a coil and a capacitor, such as those we have just described, produce varying magnetic and electric fields confined to the neighbourhood of the two components (p. 977). We now discuss how varying magnetic and electric fields can travel through space in the form of *electromagnetic waves*.

Production of Electric Field from Moving Magnetic Field

In chapter 36, we saw that a changing magnetic flux linkage produces an induced e.m.f. It was shown there that the e.m.f. induced in a rod of length l moving in a field of induction B with velocity v , when B , v and l are mutually perpendicular, is given by

$$\text{induced e.m.f.} = Blv.$$

The same e.m.f., Blv , is produced if the rod remains stationary, whilst a finite field moves with velocity v in the opposite direction to the moving

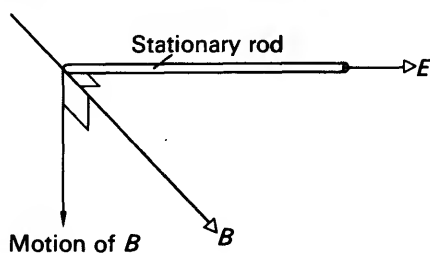


FIG. 39.15. Electric field E produced by moving B .

rod in Fig. 39.15. Hence the *electric field intensity* E , set up along the rod by the moving magnetic field is given by

$$\begin{aligned} E &= \text{potential gradient} \\ &= \frac{\text{e.m.f.}}{l} = \frac{Blv}{l} = Bv \end{aligned} \quad (1)$$

Notice that the electric field E produced is at right angles to B and v , and is in the same direction as the induced e.m.f. The directions of E , B and v are shown in Fig. 39.15.

Production of Magnetic Field from Moving Electric Field

It has already been shown that moving charges in a wire produce a magnetic field, p. 931. We can, however, look at this in a different way. We can say that each charge produces an electric field moving with the charge, and that *this moving electric field sets up a magnetic field*.

Fig. 39.16 shows a wire carrying a current due to moving electrons. The positive ions in a wire are fixed and so the electric fields of the positive ions and the negative electrons cancel. The ions are not, however, moving. Thus the magnetic field due to the moving electric

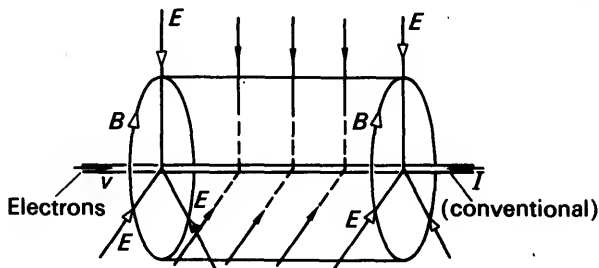


FIG. 39.16. Magnetic field B produced by moving E .

field of the electrons is *not* cancelled. Consider now a cylindrical surface of length l round the wire. If the charge per metre of the wire is ρ , then the total flux of the electric field entering the surface due to the negative charges inside = $E \times \text{area} = E \cdot 2\pi r l$, where E is the electric intensity. From Gauss's theorem (p. 747), this is equal to *charge*/ ϵ_0 or $\rho l / \epsilon_0$.

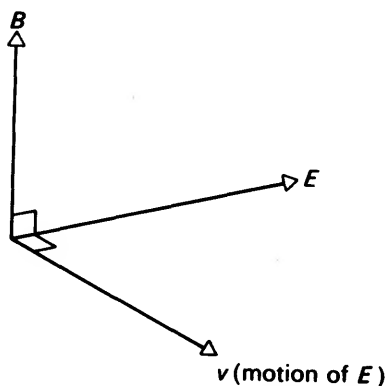
$$\therefore E = \frac{\rho}{2\pi\epsilon_0 r}.$$

But the current I = charge per second crossing a section of wire = ρv . Hence

$$E = \frac{I}{2\pi\epsilon_0 r v}.$$

From p. 934, the magnetic field produced by the current has an induction B given by $B = \mu_0 I / 2\pi r$. Substituting $I / 2\pi r = B / \mu_0$,

$$\begin{aligned} \therefore E &= \frac{B}{\epsilon_0 \mu_0 v} \\ \therefore B &= \mu_0 \epsilon_0 v E \end{aligned} \quad (2)$$

FIG. 39.17. Directions of E , B , v .

The direction of B , v and E are shown in Fig. 39.17. Note that the electric field intensity E , and the magnetic field induction B produced, are perpendicular to each other and to v .

It should also be noted that Fig. 39.17 has exactly the *same* relation between the directions of v , B and E as Fig. 39.15.

Electromagnetic Waves

Consider a pattern of electric fields E and magnetic fields B which are imagined to move into the paper with speed v as shown in Fig. 39.18.

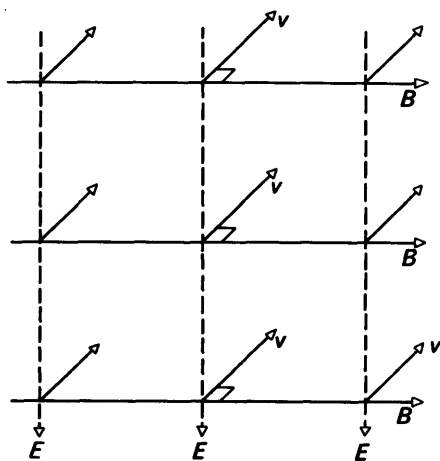


FIG. 39.18. Electromagnetic waves.

From equation (1), p. 980, the moving magnetic field creates an electric field of strength E' given by,

$$E' = vB \quad (3)$$

This electric field is in the same direction as E , as can be seen by applying Fig. 39.15. From equation (2), p. 981, the moving electric field itself creates a magnetic field of induction B' given by

$$B' = \mu_0 \epsilon_0 v E \quad . \quad . \quad . \quad (4)$$

This field is in the direction of the magnetic field B , as can be seen by applying Fig. 39.17. Now if $E = E'$ and $B = B'$, the magnetic field creates the electric field and vice versa. The system is then *self sustaining*. This occurs if

$$\begin{aligned} E' &= v B' \\ &= v(\mu_0 \epsilon_0 v E') \\ &= \mu_0 \epsilon_0 v^2 E', \end{aligned}$$

or if
$$v^2 = \frac{1}{\mu_0 \epsilon_0} \quad . \quad . \quad . \quad (5)$$

Hence if the pattern moves with a speed $1/\sqrt{\mu_0 \epsilon_0}$ the pattern will propagate through space. Now $\mu_0 = 4\pi \times 10^{-7} \text{ H m}^{-1}$ by definition, and $\epsilon_0 = 8.85 \times 10^{-12} \text{ F m}^{-1}$ by experiment.

$$\therefore v = \frac{1}{\sqrt{(4\pi \times 10^{-7} \times 8.85 \times 10^{-12})}} = 3 \times 10^8 \text{ m s}^{-1} \text{ (approx.)}$$

This is just the speed of light and radio waves, as determined by experiment. It is reasonable to suppose, therefore, that waves such as light and radio waves which move with this speed are *electromagnetic* in nature.

Demonstration of Electromagnetic Waves. Transmission Line

Electromagnetic waves of high frequency, such as radio waves, can be radiated into space by *aerials*. The electrical circuit producing the waves is called an *oscillator* (p. 1019) and it is connected to the aerial by wires called a *transmission line*.

The effect of electromagnetic waves travelling in a transmission line can be demonstrated with the aid of a very high frequency (v.h.f.) oscillator producing electrical oscillations of about 150 MHz frequency. The output from this oscillator is fed to two parallel transmission lines a few centimetres apart. Fig. 39.19. The line is terminated by a short circuit (that is, a piece of wire clipped to the two lines),

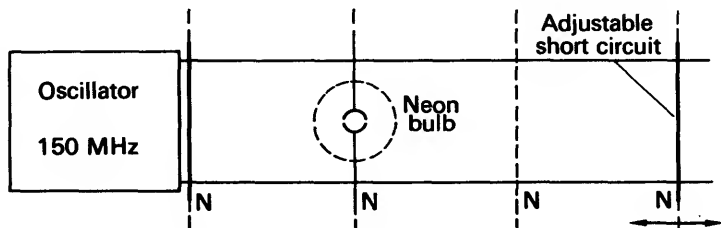


FIG. 39.19. Transmission line. Neon detector.

whose position along the line may be varied. As a neon bulb is moved along the line, variations in intensity are produced; the peaks of intensity are separated by a distance of about 100 cm. The greatest variations are obtained by adjusting the position of the short circuit. At the points marked by N in the diagram the neon may then be extinguished; mid-way between these positions the neon is brightest.

At the points marked N, therefore, there is no voltage and these points are *voltage nodes* (compare *sound waves*, p. 642). Mid-way between consecutive nodes, the high frequency p.d. has a maximum value. At these points on the wire, therefore, there are *voltage antinodes*. Thus a *voltage standing wave* is produced on the line. Since nodes are separated by half a wavelength (p. 590), the wavelength λ can be measured from this experiment; and if the oscillator frequency f is known, the speed c of the waves can be found from the general relation $c = f\lambda$. A value of about $3 \times 10^8 \text{ m s}^{-1}$ is obtained for electromagnetic waves.

If the neon bulb is replaced by a small filament lamp (which, unlike the neon, has a very low resistance) it is found that the lamp glows brightest, not at the voltage antinodes, but at the voltage nodes. This is because the low resistance lamp is sensitive to *current* in the line and not to voltage. Thus there are current antinodes at voltage nodes, and current nodes at voltage antinodes, as illustrated in Fig. 39.20.

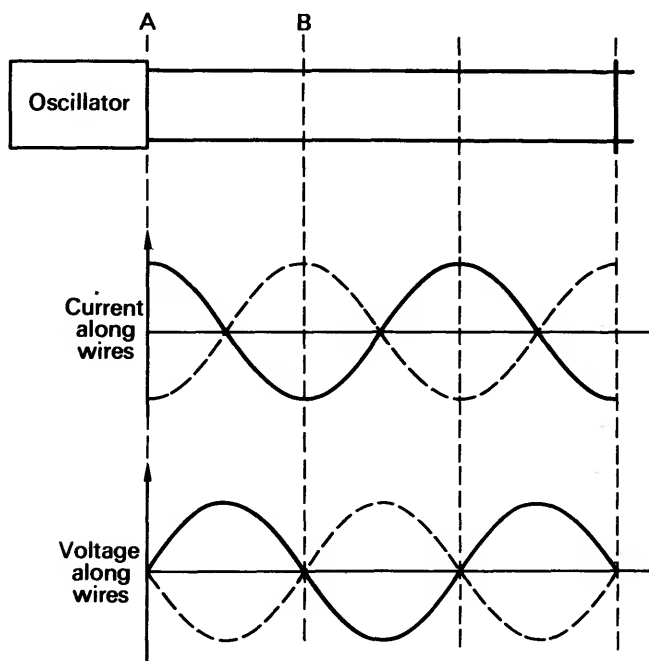


FIG. 39.20. Standing waves of voltage and current.

This situation is very like the standing waves set up in a resonating pipe open at both ends (p. 648). In this electrical case, resonance is obtained by adjusting the short circuit. In a pipe, resonance may be obtained by adjusting the length of the tube. As illustrated in Fig. 39.31,

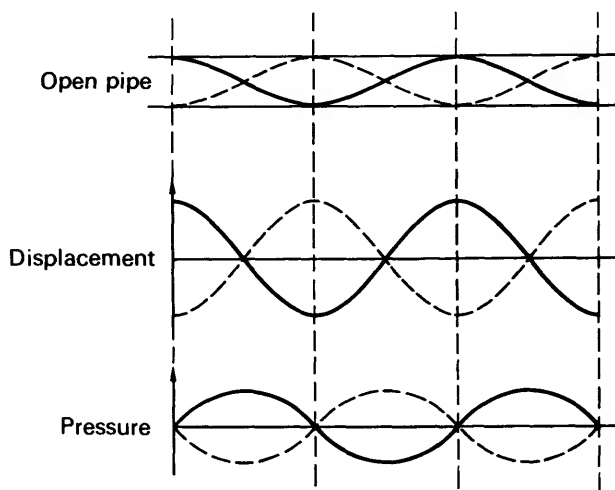


FIG. 39.21. Standing sound waves.

pressure nodes in a resonating pipe exist at displacement antinodes, and vice-versa.

To see what is happening on the transmission line during the course of one oscillation of the oscillator, consider the section of the line from A to B (Fig. 39.22). As the oscillator begins to force charge from the lower wire to the upper wire, a current will be flowing past A to the right on the upper wire, and to the left on the lower wire. Current will also be flowing past B but to the left on the upper wire and to the right on the lower wire, as shown in Fig. 39.22.

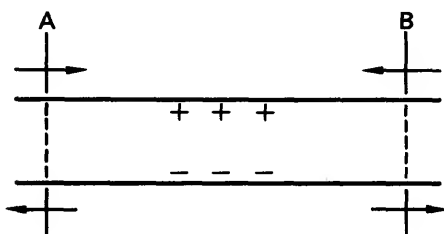


FIG. 39.22. Charge movement on transmission line.

This means that positive charge accumulates between A and B on the upper wire, so that the centre point will increase in potential. On the lower wire there will be a deficit of charge. Thus the mid-point of AB on the lower wire will be at a low potential. It follows that the p.d. across the wire at the mid-point increases. This will continue for one quarter of a cycle, when the process starts to reverse and the directions of current flow change. Later in the cycle, there will be an

accumulation of positive charge on the lower wire. The process is then repeated and a standing wave pattern on the wire is therefore produced. The moving charges create electric and magnetic fields between the wires which constitute an electromagnetic wave. Here the wave is a standing wave; it is due to two electromagnetic waves travelling in opposite directions.

Radiation of Electromagnetic Waves into Space

Consider an oscillator with a connected transmission line, and suppose the transmission line is bent as shown in Fig. 39.23 (i).

We choose the points C and D to be at a current antinode, and A and B at a current node. It is clear that no current can flow at A and B as the

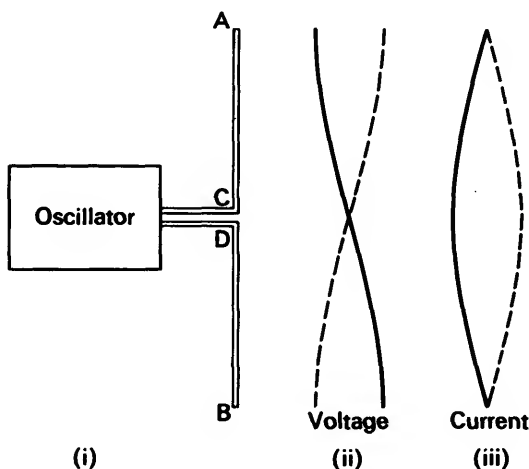


FIG. 39.23. Aerial. Half-wave dipole.

wires terminate at these points. A standing wave pattern will be set up on the wires, as we have seen, and if the lengths AC and DB are chosen correctly, there will be no additional nodes on the wires. In this case AB must be half a wavelength. Such an arrangement acts as an *aerial*, as discussed shortly. It is called a *half-wave dipole aerial*.

The charges moving along AB are forced up to A during one half cycle of oscillation and then down to B during the next half cycle. The charge, therefore, oscillates between A and B. *This accelerating charge radiates energy in the form of electromagnetic waves.* In contrast, charges moving in a line with a steady speed create a static magnetic field, and no electromagnetic wave is radiated.

If this aerial is used then, electromagnetic waves are radiated into space. They may be detected by a similar receiving aerial, of the same length as the transmitting aerial. Fig. 39.24. Detection of the signal is obtained using a wireless receiver tuned to the correct frequency.

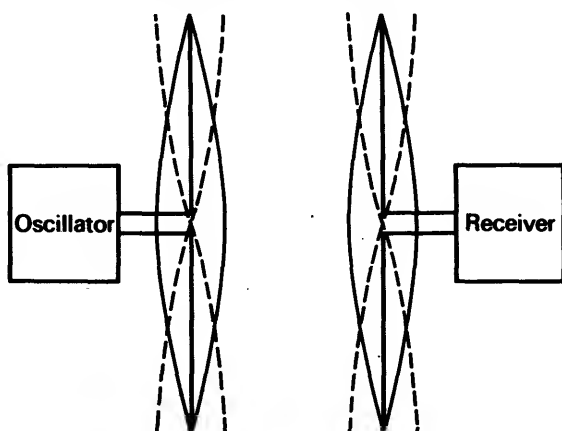


FIG. 39.24. Transmitting and receiving aerials.

Polarization of Electromagnetic Waves

By rotating the receiving aerial about the axis AB, it can be shown that no signal is obtained in the receiver when the aerial is directed along the line MN. Fig. 39.25. This corresponds to an angle of rotation α of 90° . Thus the radiation from the transmitting aerial is *plane polarized*, that is, the electric and magnetic vibrations in the electromagnetic wave each occur in one plane. The distribution of electric and

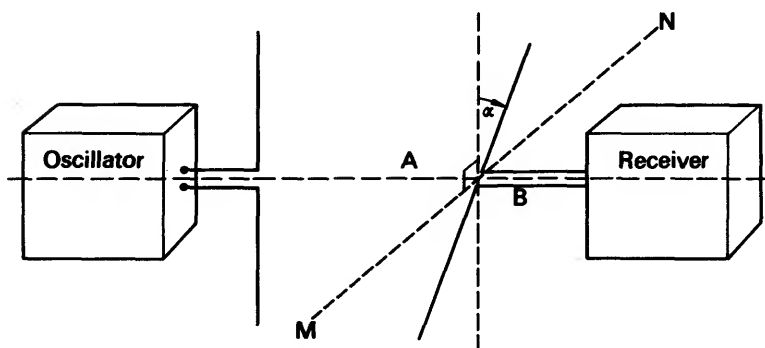


FIG. 39.25. Polarization of electromagnetic wave.

magnetic fields in the plane-polarized wave is shown in Fig. 39.26. The plane containing the magnetic field B is at right angles to the plane containing the electric field E , and the directions of E , B and v are mutually perpendicular (compare Fig. 39.15 and 39.17).

As we have seen, light is an example of an electromagnetic wave. The phenomenon of polarization of light is explained in an exactly

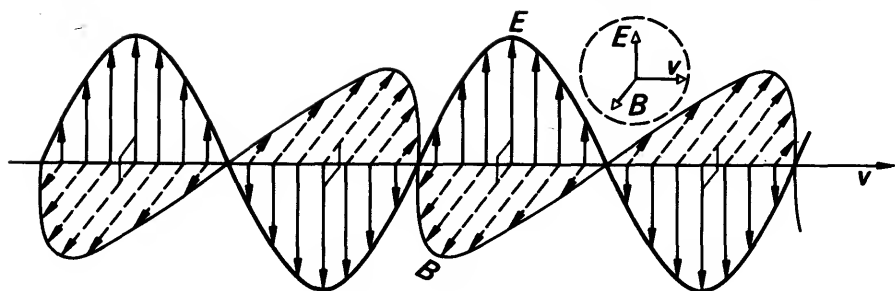


FIG. 39.26. Plane-polarized wave.

similar way (see p. 716). Plane polarized light has its electric field in one plane and its magnetic field in a perpendicular plane. Unpolarized light has electric and magnetic vibrations in *all* planes perpendicular to the direction of propagation.

Wave Properties of Electromagnetic Waves

We have already seen how standing (stationary) waves on transmission lines (guided waves), and the polarization of electromagnetic waves, can be demonstrated. It can also be shown that electromagnetic waves, like all waves, can undergo reflection, refraction, interference and diffraction. The apparatus described on p. 987 has a frequency of 150 MHz and hence the wavelength $\lambda = c/f = 3 \times 10^8 / (150 \times 10^6)$ metres = 2 metres = 200 cm. This is rather long for demonstration experiments in a teaching laboratory. *Microwaves* of about 3 cm wavelength are therefore used. These are radiated from a horn waveguide T and are received by a similar waveguide R or smaller *probe*. The detected wave then produces a deflection in a connected meter. Some experiments which can be performed in a school laboratory are illustrated in Fig. 39.27 (i)–(vi).

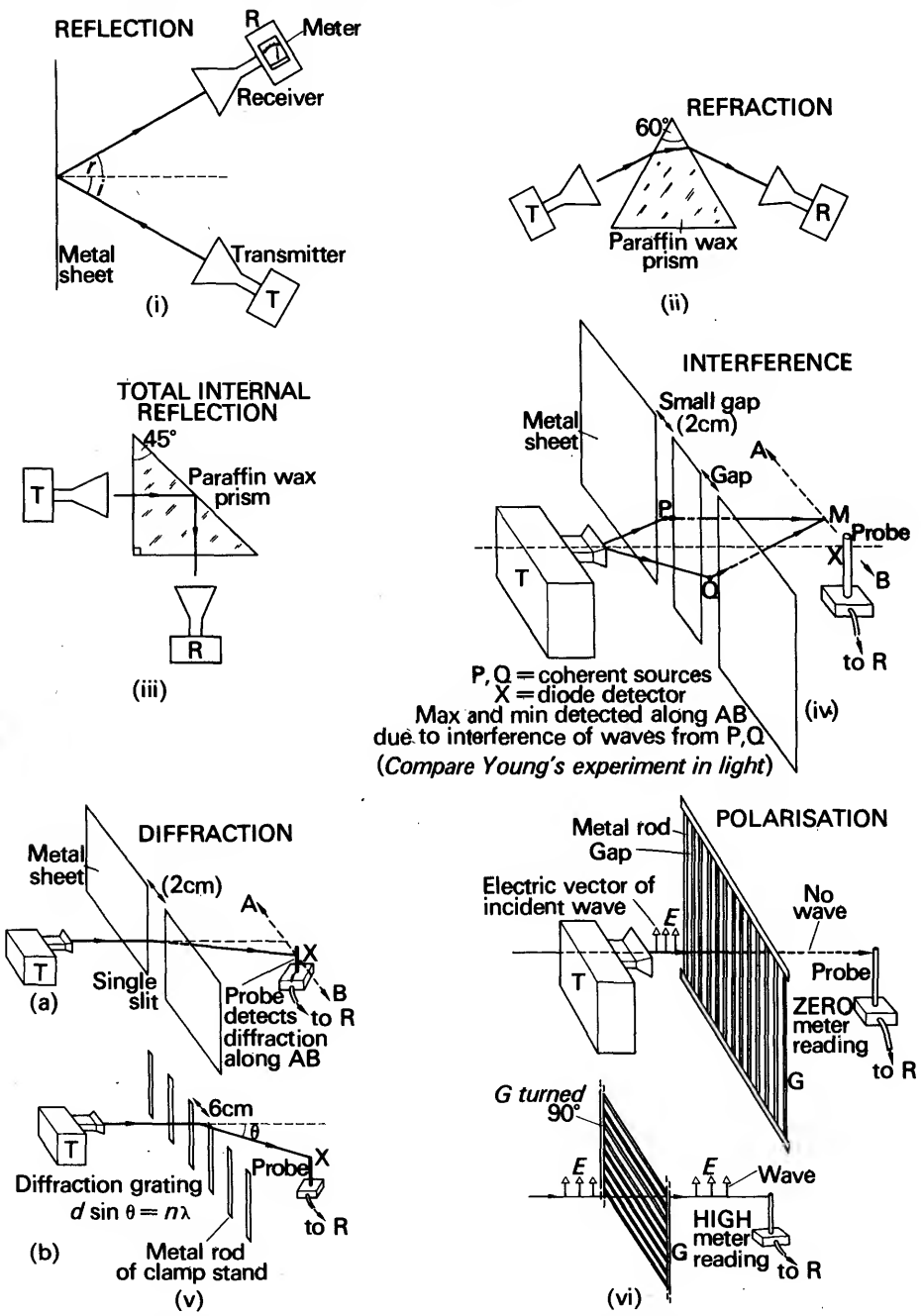


FIG. 39.27. Experiments with microwaves: (i) Reflection; (ii) Refraction; (iii) Total internal reflection; (iv) Interference; (v) Diffraction; (vi) Polarization.

EXERCISES 39

1. Explain what is meant by the *peak value* and *root mean square value* of an alternating current. Establish the relation between these quantities for a sinusoidal waveform.

What is the r.m.s. value of the alternating current which must pass through a resistor immersed in oil in a calorimeter so that the initial rate of rise of temperature of the oil is three times that produced when a direct current of 2 amps passes through the resistor under the same conditions? (*N.*)

2. Define the *impedance* of an a.c. circuit.

A $2.5 \mu\text{F}$ capacitor is connected in series with a non-inductive resistor of 300 ohms across a source of p.d. of r.m.s. value 50 volts alternating at $1000/2\pi$ Hz. Calculate (*a*) the r.m.s. values of the current in the circuit and the p.d. across the capacitor, (*b*) the mean rate at which energy is supplied by the source. (*N.*)

3. Describe and explain the mode of action of an ammeter suitable for the measurement of alternating current.

A constant a.c. supply is connected to a series circuit consisting of a resistance of 300 ohms in series with a capacitance of 6.67 microfarads, the frequency of the supply being $3000/2\pi$ Hz. It is desired to reduce the current in the circuit to half its value. Show how this could be done by placing either (*a*) an additional resistance, or (*b*) an inductance, in series. Calculate in each case the magnitude of the extra component. (*L.*)

4. Describe and explain an instrument for measuring alternating current. What do you understand by the r.m.s. value of an alternating current? How is it related to the peak value in the case of a sinusoidal current?

When the coils of an electromagnet are connected to a 240-volt d.c. supply, the current taken is 10 amps. When connected to a 240-volt a.c. supply the current taken is only 1 amp. Explain why there is a smaller current on a.c. and calculate the resistance of the coils. Using the same time axis, draw curves showing how the current through the coils and the applied voltage vary with time when the supply is a.c. (*C.*)

5. A rectangular coil with n turns each of area A is rotated with uniform angular velocity ω about an axis at right angles to a uniform magnetic field of flux density B . Show that a sinusoidal alternating e.m.f. is generated in the coil and write down an expression for its peak value.

Explain what is meant by the root-mean-square (r.m.s.) value of an alternating current or voltage, and show how for a sinusoidal alternation it can be calculated from the peak value.

A resistor and an inductor are connected in series across the output terminals of a variable frequency a.c. generator. The r.m.s. voltage between the terminals of the generator is 10 volts. When the frequency is 20 cycles per second, the r.m.s. current in the circuit is 0.20 amp, the p.d. across the resistor is 8 volts, and the p.d. across the inductor is 6 volts. With the help of a vector diagram, explain these readings, and find the values of the resistance and the inductance.

With the generator voltage unchanged at 10 volts r.m.s. its frequency is adjusted until at some value f the p.d. across the resistor and that across the inductor both have the same value V . Calculate f and V . (*O.*)

6. Define *self inductance*. Explain and justify a method of calculating the self inductance of an endless solenoid whose axis is a circle.

A coil of self inductance 0.987 millihenry and resistance 25 ohms is connected in series with a variable capacitance and an a.c. supply of constant e.m.f. and of frequency $15000/\pi$ Hz. Find the value of the capacitance for which the current

in the coil would be a maximum. Represent on a vector diagram the relative potential differences across the coil and the capacitance in this condition and find the phase angle between them. What is the power factor of the coil? (L.)

7. A box X and a coil Y are connected in series with a variable frequency a.c. supply of constant e.m.f. 10 volts. X contains a capacitance $1\ \mu\text{F}$ in series with a resistance 32 ohms, and Y has a self-inductance 5.1 millihenry and resistance 68 ohms. The frequency is adjusted until maximum current flows in X and Y . Determine the impedances of X and Y at this frequency and hence find the potential differences across X and Y separately. Show these potential differences on a vector diagram indicating their phase relations with the applied e.m.f. (L.)

8. Explain what is meant by *self-induction*. Describe one experiment in each case to illustrate the effect of a coil having self-inductance in (a) a direct current circuit, (b) an alternating current circuit.

Describe the essential features of an instrument for measuring alternating current. (N.)

9. A filament lamp of negligible inductance is found to glow when connected in series with a capacitor across an alternating voltage supply of negligible impedance. Explain what will happen to the brightness of the lamp if the frequency of the supply is doubled. Show how a second capacitor can be included in the circuit so as to increase the brightness of the lamp.

What is the advantage of using a capacitor rather than a resistor to control the brightness of a lamp operated on an alternating voltage power supply? (N.)

10. Explain what is meant by the terms *impedance* and *phase angle* as applied to a circuit carrying alternating current.

An a.c. generator of negligible internal impedance whose output voltage is 100 volts r.m.s. is connected in series with a resistance of 100 ohms and a pure inductance of 2 henrys. If an a.c. voltmeter gives the same reading whether connected across the resistance or the inductance what does it read? What is the frequency of the generator? Calculate the power dissipated in the resistor at the instant when the current is half its maximum value. Calculate also the power supplied to the inductance at the same instant. Show that the sum of these two quantities is equal to the power delivered by the generator at this instant. (O. & C.)

11. An alternating current is represented by the equation $I = I_0 \sin \omega t$. Write down (a) the r.m.s. value of the current, (b) the frequency of the current. Deduce the maximum value of the rate of change of the current.

A long solenoid wound with 10 turns per cm carries an alternating current of frequency 50 cycles per second⁻¹ and of r.m.s. value 1 amp. In the centre of the solenoid is mounted coaxially a five-turn coil of diameter 1 cm. Calculate the r.m.s. voltage induced in this coil.

If the inner coil is surrounded by a copper tube, the voltage induced in it decreases. Explain. (C.)

12. The r.m.s. potential difference V volts alternating at f Hz across a coil is adjusted until the r.m.s. current through the coil is 10 mA. Corresponding values of V and f are given in the table, the current being adjusted to 10 mA in each instance.

f (Hz)	1000	1500	2000	2500	3000
V (volts)	3.88	5.78	7.58	9.46	11.30

(a) Why does V change when f changes? (b) Plot a graph of V^2 against f^2 . (c) Use the graph to find the inductance of the coil. (d) Comment on the possible use of the graph in a determination of the resistance of the coil. (e) If a corresponding set of readings were obtained using a capacitor instead of a coil, state with reasons how V would vary with f . (N.)

13. State the laws of electromagnetic induction and describe briefly experiments to show their validity.

A coil A passes current of 1.25 A when a steady potential difference of 5 V is maintained across it, and an r.m.s. current of 1 A when it has across it a sinusoidal potential difference of 5 V r.m.s. at a frequency of 50 Hz (cycles per second). Explain why the current is less in the second case, and calculate the resistance and the inductance of the coil.

The same coil A , which has 100 turns, has a second coil B with 500 turns wound on it so that all the magnetic flux produced by A is linked by B . Find the r.m.s. value of the e.m.f. that appears across the open-circuit ends of B when a sinusoidal alternating current of 1 A r.m.s. at a frequency of 50 Hz is passed through A . Why is the ratio of this e.m.f. to the r.m.s. potential difference across A not the same as the ratio of the number of turns in B and A , i.e. 5:1?

Explain why the insertion of an iron core into the coils would decrease the current in A and increase the e.m.f. across B , if the alternating potential difference across A were kept unchanged; the effects of hysteresis and eddy currents in the iron may be neglected. (O. & C.)

Electromagnetic waves

14. A uniform magnetic field of 0.5 Wb m^{-2} is moving with a constant velocity of 4 m s^{-1} past a stationary straight conductor of length 20 cm. If the field, velocity and conductor are mutually perpendicular, calculate the electric field intensity in the conductor observed by an observer moving with the field and the force on an electron in the conductor ($e = 1.6 \times 10^{-19} \text{ C}$).

15. A wire in the form of a circle radius 5 cm carries a charge of 10^{-8} C and rotates with a constant speed of $10 \text{ rev. second}^{-1}$ about its centre. Considering the moving charge as a current, calculate the magnetic field at the centre of the field.

16. 'A moving electric field E produces a magnetic field B .' Explain this statement and derive a relation for B .

17. Show that a magnetic field B , moving with uniform velocity v , generates an electric field $E = Bv$.

18. Stationary electromagnetic waves can be set up in a transmission line. Explain, with diagrams, how this can be demonstrated in the laboratory and how you would verify the existence of the stationary waves.

19. The length of a half wave dipole is 15 m and the frequency of the oscillator feeding the aerial is increased slowly from 5 MHz to 15 MHz. State and explain what will be observed on a receiver connected to an exactly similar aerial, which is placed parallel to the first aerial and a short distance away.

How would you show that the electromagnetic waves from the dipole are plane polarized?

chapter forty

Electrons. Motion in Fields

Particle Nature of Matter

MATTER is made up of many millions of molecules, which are particles whose dimensions are about 3×10^{-10} m. Evidence of the existence of molecules is given by experiments demonstrating *Brownian motion*, with which we assume the reader is familiar. One example is the irregular motion of smoke particles in air, which can be observed by means of a microscope. This is due to continuous bombardment of a tiny smoke particle by numerous air molecules all round it. The air molecules move with different velocities in different directions. The resultant force on the smoke particle is therefore unbalanced, and irregular in magnitude and direction. Larger particles do not show Brownian motion when struck on all sides by air molecules. The resultant force is then relatively negligible.

More evidence of the existence of molecules is supplied by the successful predictions made by the *kinetic theory of gases*. This theory assumes that a gas consists of millions of separate particles or molecules moving about in all directions. *X-ray diffraction patterns* of crystals also provides evidence for the particle nature of matter. The symmetrical patterns of spots obtained are those which one would expect from a three-dimensional grating or lattice formed from particles. A smooth continuous medium would not give a diffraction pattern of spots.

The size of atoms and molecules can be estimated in several different ways. By allowing an oil drop to spread on water, for example, an upper limit of about 5×10^{-7} cm is obtained for the size of an oil molecule. X-ray diffraction experiments enable the interatomic spacing between atoms in a crystal to be accurately found. The results are of the order of a few angstrom units, such as 3 Å or 3×10^{-10} m or 0.3 nm (nanometre).

A simple calculation shows the order of magnitude of the enormous number of molecules present in a small volume. One gram of water occupies 1 cm³. One mole has a mass of 18 g, and thus occupies a volume of 18 cm³. Assuming the diameter of a molecule is 3×10^{-8} cm, its volume is roughly $(3 \times 10^{-8})^3$ or 27×10^{-24} cm³. Hence the number of molecules in 18 cm³ = $18/(27 \times 10^{-24}) = 6 \times 10^{23}$ approximately.

Avogadro's constant, N_A , is the number of molecules in one mole of a substance. Accurate values show that $N_A = 6.02 \times 10^{23} \text{ mol}^{-1}$, or $6.02 \times 10^{26} \text{ kmol}^{-1}$, where 'kmol' represents a kilomole.

Particle Nature of Electricity

In electrolysis, we assume that the carriers of current through an acid or salt solution are ions, which may be positively and negatively charged (p. 846). From Faraday's laws of electrolysis, the charge carried by each ion is proportional to its valency (p. 845). We can find the charge on a monovalent ion using the following argument.

Avogadro's constant, about 6.02×10^{23} , is the number of molecules in one mole. In electrolysis, 96500 coulombs (one faraday) is the quantity of electricity required to deposit one mole of a monovalent element (see p. 845). When the element is monatomic, the number of ions of one kind which carry this charge is equal to the number of molecules. Thus the charge on each ion is given by $96500/6.02 \times 10^{23}$ or $1.6 \times 10^{-19} \text{C}$. If $1.6 \times 10^{-19} \text{C}$ is denoted by the symbol e , the charge on any ion is then e , $2e$, or $3e$, etc., depending on its valency. Thus e is a basic unit of charge.

All charges, whether produced in electrostatics, current electricity or any other method, are multiples of the basic unit e . Evidence that this is the case was obtained by Millikan, who, in 1909, designed an experiment to measure the unit e .

Theory of Millikan's experiment

Millikan first measured the terminal velocity of an oil-drop falling through air. He then charged the oil-drop and applied an electric field to oppose gravity. The drop now moved with a different terminal velocity, which was again measured.

Suppose the radius of the oil-drop is a , the densities of oil and air are ρ and σ respectively, and the viscosity of air is η . When the drop, without a charge, falls steadily under gravity with a terminal velocity v_1 ,

upthrust + viscous force = weight of drop.

$$\therefore \frac{4}{3}\pi a^3 \sigma g + 6\pi\eta a v_1 \text{ (Stokes's law)} = \frac{4}{3}\pi a^3 \rho g,$$

$$\therefore 6\pi\eta a v_1 = \frac{4}{3}\pi a^3 (\rho - \sigma)g, \quad \quad \quad (i)$$

$$\therefore a = \left[\frac{9\eta v_1}{2(\rho - \sigma)g} \right]^{\frac{1}{2}} \quad \quad \quad (ii)$$

Suppose the drop now acquires a negative charge e' and an electric field of intensity E is applied to oppose gravity, so that the drop now has a terminal velocity v_2 . Then, since the force due to E on the drop is Ee' , we have

$$\frac{4}{3}\pi a^3 \sigma g + 6\pi\eta a v_2 = \frac{4}{3}\pi a^3 \rho g - Ee'.$$

$$\therefore Ee' = \frac{4}{3}\pi a^3 (\rho - \sigma)g - 6\pi\eta a v_2.$$

Hence, from (i), $Ee' = 6\pi\eta a v_1 - 6\pi\eta a v_2 = 6\pi\eta a (v_1 - v_2).$

Thus, with (ii) $e' = \frac{6\pi\eta \left[\frac{9\eta v_1}{2(\rho - \sigma)g} \right]^{\frac{1}{2}} (v_1 - v_2)}{E} \quad \quad \quad (iii)$

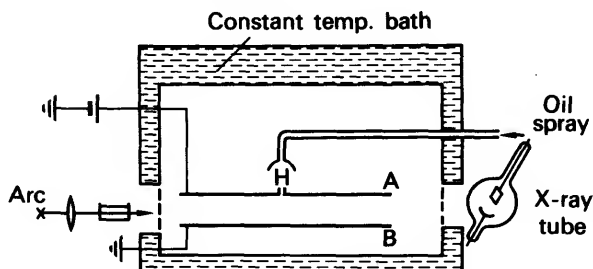


FIG. 40.1. Principle of Millikan's experiment.

Experiment

In his experiments Millikan used two horizontal plates A, B about 20 cm in diameter and 1.5 cm apart, with a small hole H in the centre of the upper plate (Fig. 40.1). He used a fine spray to 'atomize' the oil and create tiny drops above H, and occasionally one would find its way through H, and would be observed in a low-power microscope by reflected light when the chamber was brightly illuminated. The drop was seen as a pin-point of light, and its downward velocity was measured by timing its fall through a known distance by means of a scale in the eyepiece. The field was applied by connecting a battery of several thousand volts across the plates A, B, and its intensity E was known, since $E = V/d$, where V is the p.d. between the plates and d is their distance apart. Millikan found that the friction between the drops when they were formed by the spray created electric charge, but to give a drop an increased charge an X-ray tube was operated near the chamber to ionize the air.

From equation (iii), it follows that when v_1 , v_2 , E , ρ , σ and η are all known, the charge e' on the drop can be calculated. Millikan found, working with hundreds of drops, that the charge e' was always a simple multiple of a basic unit, 1.6×10^{-19} coulomb. He thus concluded that the charge e was numerically 1.6×10^{-19} coulomb.

EXAMPLE

Calculate the radius of a drop of oil, density 900 kg m^{-3} , which falls with a terminal velocity of $2.9 \times 10^{-2} \text{ cm s}^{-1}$ through air of viscosity $1.8 \times 10^{-5} \text{ N s m}^{-2}$. Ignore the density of the air.

If the charge on the drop is $-3e$, what p.d. must be applied between two plates 5 cm apart for the drop to be held stationary between them? ($e = 1.6 \times 10^{-19} \text{ C}$.)

When the drop falls with a terminal velocity, force due to viscous drag = weight of sphere. With the usual notation, if ρ is the oil density, we have

$$6\pi\eta av = \text{volume} \times \text{density} \times g = \frac{4}{3}\pi a^3 \rho g$$

$$\therefore a = \sqrt{\frac{9\eta v}{2\rho g}} = \sqrt{\frac{9 \times 1.8 \times 10^{-5} \times 2.9 \times 10^{-4}}{2 \times 900 \times 9.8}}$$

$$= 1.6 \times 10^{-6} \text{ m} = 1.6 \times 10^{-4} \text{ cm}, \quad (1)$$

since $v = 2.9 \times 10^{-2} \text{ cm s}^{-1} = 2.9 \times 10^{-4} \text{ m s}^{-1}$ and $g = 9.8 \text{ m s}^{-2}$.

Suppose the upper plate is V volts higher than the lower plate when the drop is stationary, so that the electric field intensity E between the plates is V/d . Then upward force on drop $= E \times 3e =$ weight of drop.

$$\therefore E \times 3e = \frac{4}{3}\pi a^3 \rho$$

$$\therefore E = \frac{4\pi a^3 \rho}{9e} = \frac{V}{d}$$

$$\therefore V = \frac{4\pi a^3 \rho d}{9e}$$

$$= \frac{4\pi \times (1.6 \times 10^{-6})^3 \times 900 \times 5 \times 10^{-2}}{9 \times 1.6 \times 10^{-19}}$$

$$= 1600 \text{ V.}$$

CATHODE RAYS (ELECTRONS) AND PROPERTIES

Atomic physics can be said to have begun with the study of the conduction of electricity through gases. The passage of electricity through a gas, called a 'discharge', was familiar to Faraday, but the steady conduction—as distinct from sparks—takes place when the pressure of the gas is less than about 50 mm Hg; in a neon lamp it is about 10 mm Hg.

The Gaseous Discharge at Various Pressures

Fig. 40.2 (i) represents a glass tube, about 0.5 metre long, connected to a vacuum pump P and a pressure gauge G . It contains an anode A and a cathode K , connected respectively to the positive and negative terminals of the secondary of an induction coil. As the air is pumped out, nothing happens until the pressure has fallen to about 100 mm Hg (mercury). Then thin streamers of luminous gas appear between the electrodes (Fig. 40.2 (ii)).

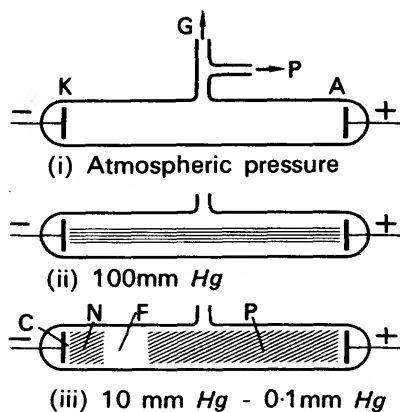


FIG. 40.2. Stages in development of gaseous discharge.

At about 10 mm Hg the discharge becomes a steady glow, spread throughout the tube (Fig. 40.2 (iii)). It is broken up by two darker regions, of which the one nearest the cathode, K in the figure, is narrow and hard to see. The dark region C is called the cathode dark space, or sometimes, after its discoverer, the Crookes' dark space. Beyond the cathode dark space is a bright region N called the negative glow, and beyond that the Faraday dark space F—also called after its discoverer. Beyond the Faraday dark space stretches a luminous column P, called the positive column, which fills the rest of the discharge tube. Sometimes the positive column breaks up into alternating bright and dark segments, called striations, shown in Fig. 40.3. In all the photographs

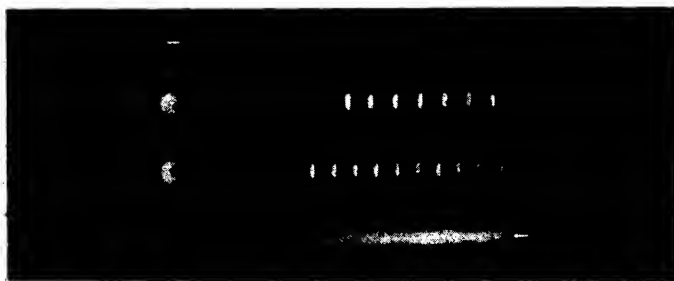


FIG. 40.3.

The lowest photograph shows the positive column. As the pressure decreases, the positive column breaks up into striations and shrinks towards the anode on the right. The top photograph shows the dark space completely filling the tube, when the pressure is about 0.01 mm Hg.

of Fig. 40.3 the cathode is on the left. The cathode dark space can hardly be seen—it lies just around the cathode—but the negative glow and Faraday dark space are clear.

The positive column is the most striking part of the discharge, but the cathode dark space is electrically the most important. In it the electrons from the cathode are being violently accelerated by the electric field, and gaining energy with which to ionize the gas atoms. In the positive column some atoms are being ionized by collisions with electrons; others are being excited, in a way which we cannot describe here, and made to emit their characteristic spectra.

When the pressure of the gas in the discharge tube is reduced still further, the dark spaces swell, and the positive column shrinks. At about 1 mm Hg the cathode dark space becomes distinct, and at 0.1 mm Hg it is several centimetres long. Eventually, as the pressure falls, the cathode dark space stretches from the anode to the cathode, and the negative glow and positive column vanish. This happens at about 0.01 mm Hg in a tube about a half-metre long.

When the cathode dark space occupies the whole discharge tube, the

walls of the tube fluoresce, in the way we have already described. The electrons flying across the space are called *cathode rays*. Where they strike the anode they produce X-rays (p. 1067).

The Mechanism of Conduction

How are the ions and electrons in a gaseous discharge produced? A luminous discharge requires a voltage, V , of at least a hundred volts across the gas at pressures of about 1 mm Hg. At much higher or lower pressures, it may require thousands of volts. But with a voltage of about ten, although there is no glow, a very weak current, I , can be detected—of the order of 10^{-15} amp (Fig. 40.4 (i)). This we attribute

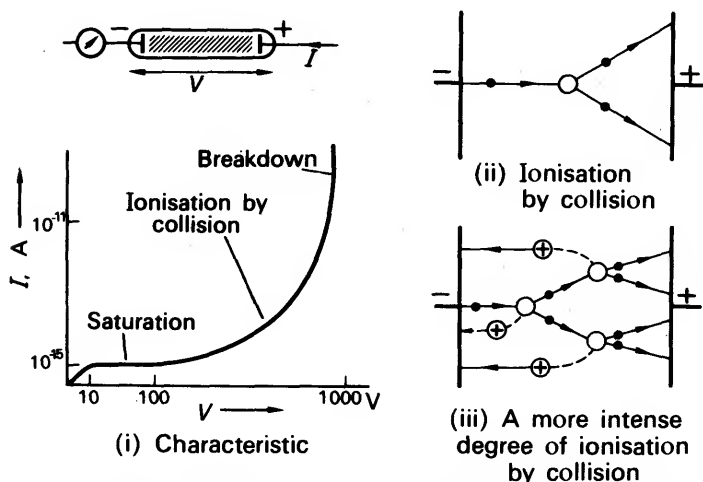


FIG. 40.4. Discharge through gas at low pressure. [In (i) currents and voltages are order of magnitude only: in (ii) and (iii), the numerous non-ionizing collisions are not shown.]

to electrons emitted from the cathode by the photo-electric effect (p. 1076); a trace of ultra-violet light in the laboratory would account for the emission. When the voltage is increased the electrons are accelerated by the electric field to a higher speed, and strike the gas atoms more violently on their way to the anode. When the voltage is high enough the electrons strike the atoms with sufficient kinetic energy to knock electrons out of them (Fig. 40.4 (ii)). This process is called ionization by collision; the atoms become ions, and move towards the cathode; the extra electrons join the original ones in their flight to the anode. At higher voltages the knocked-out electrons are accelerated enough to produce more ions and electrons on the way (Fig. 40.4 (iii)). Eventually a point is reached at which the current grows uncontrollably—the gas is said to break down. In practice, the current is limited by a resistor, in series with the discharge tube; in a commercial neon lamp this resistor, of resistance about 5000 ohms, is hidden in the base.

The current through a gas, like that through an electrolyte, is carried by carriers of both signs—positive and negative. At the anode, the

negative electrons enter the wires of the outside circuit, and eventually come round to the cathode. There they meet positive ions, which they now enter, and so re-form neutral gas atoms. Positive ions arriving at the cathode knock off some of the atoms, which diffuse into the body of the discharge, and there, eventually, they are ionized again. Thus a limited amount of gas can carry a current indefinitely.

Once a gas has broken down, current can continue to pass through it even in the dark: that is to say, when there is no ultra-violet light to make the cathode emit electrons. The electrons from the cathode are now simply knocked out of it by the violent bombardment of the positive ions.

Ultra-violet light is not, as a rule, necessary even for starting a gaseous discharge. The somewhat mysterious cosmic rays, which reach the earth from outer space, are able to ionize a gas; they may therefore enable a discharge to start. Once it has started, the emission of electrons by bombardment of the cathode keeps it going.

Modern Production of Cathode Rays

The discharge tube method is not a convenient one for producing and studying cathode rays or electrons. Firstly, a gas is needed at the appropriate low pressure; secondly, a very high p.d. is needed across the tube; thirdly, X-rays are produced (p. 1068) which may be dangerous.

Nowadays a *hot cathode* is used to produce a supply of electrons. This may consist of a fine tungsten wire, which is heated to a high temperature when a low voltage source of 4–6 V is connected to it. Metals contain free electrons, moving about rather like the molecules in a gas. If the temperature of the metal is raised, the thermal velocities of the electrons will be increased. The chance of electrons escaping from the attraction of the positive ions, fixed in the lattice, will then also be raised. Thus by heating a metal such as tungsten to a high temperature, electrons can be 'boiled off'. This is called *thermionic emission* (see also p. 1009).

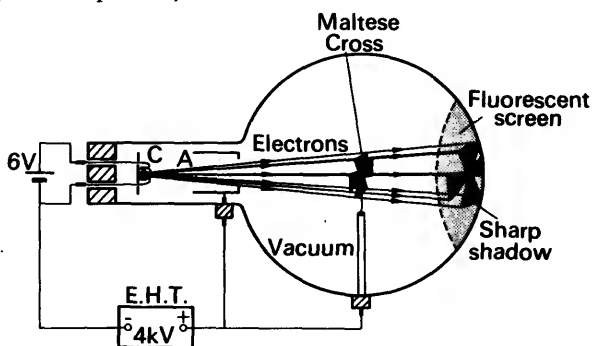


FIG. 40.5. Electrons travel in straight lines.

Fig. 40.5 shows a tungsten filament C inside an evacuated tube. When heated by a low voltage supply, electrons are produced, and they are accelerated by a positive voltage of several thousand volts

applied between C and a metal cylinder A. The electrons travel unimpeded across the tube past A, and produce a glow when they collide with a fluorescent screen and give up their energy.

Properties of Cathode Rays

Fast-moving electrons emitted from C produce a sharp shadow of a Maltese cross on the fluorescent screen, as shown in Fig. 40.5. Thus the cathode rays travel in straight lines. They also produce heat when incident on a metal—a fine piece of platinum glows, for example.

When a magnet is brought near to the electron beam, the glow on the fluorescent screen moves, Fig. 40.6. If Fleming's left hand rule is

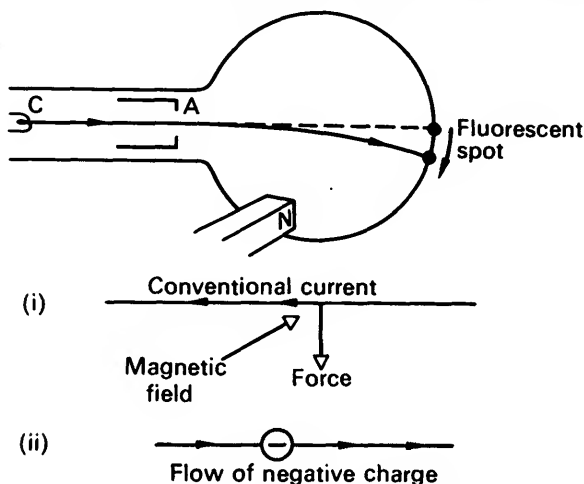


FIG. 40.6. Deflection shows electrons are negatively charged.

applied to the motion, the middle finger points in a direction *opposite* to the electron flow. Thus electrons appear to be particles which carry a *negative* charge.

This is confirmed by collecting electrons inside a *Perrin tube*, shown in Fig. 40.7. The electrons are deflected by the magnet S until they

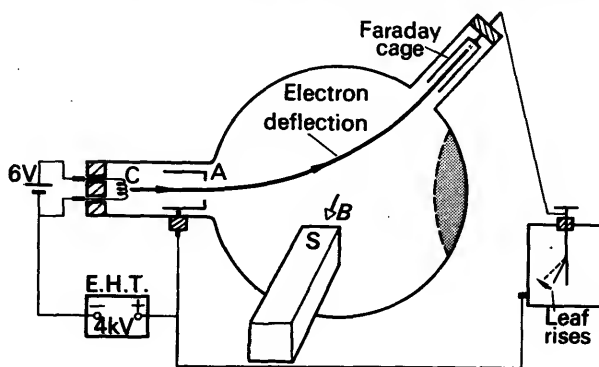


FIG. 40.7. Perrin tube. Direct method for testing electron charge.

pass into a metal cylinder called a 'Faraday cage' (see p. 760). The cylinder is connected to the plate of an electroscope, which has been negatively charged using an ebonite rod and fur. As soon as the electrons are deflected into the cage the leaf rises further, showing that an extra *negative* charge has been collected by the cage. This supports the idea that cathode rays are fast moving electrons.

ELECTRON MOTION IN ELECTRIC AND MAGNETIC FIELDS

Deflection in an Electric Field

Suppose a horizontal beam of electrons, moving with velocity v , passes between two parallel plates, Fig. 40.8. If the p.d. between the plates is V and their distance apart is d , the field intensity $E = V/d$. Hence the force on an electron of charge e moving between the plates $= Ee = eV/d$ and is directed towards the positive plate.

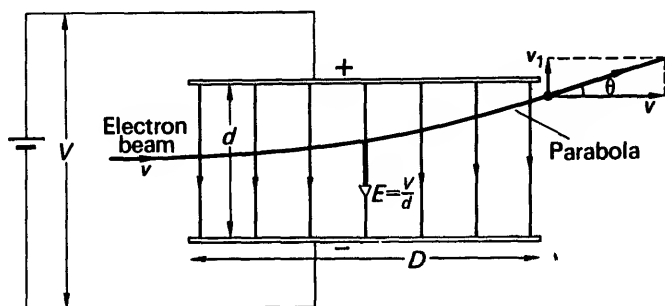


FIG. 40.8. Deflection in electric field.

Since the electric intensity E is vertical, no horizontal force acts on the electron entering the plates. Thus the horizontal velocity, v , of the beam is unaffected. This is similar to the motion of a projectile projected horizontally under gravity. The vertical acceleration due to gravity does not affect the horizontal motion.

In a vertical direction the displacement, y , $= \frac{1}{2}at^2$, where $a = \text{acceleration} = \text{force/mass} = Ee/m_e$ and t is the time.

$$\therefore y = \frac{1}{2} \frac{Ee}{m_e} t^2 \quad \dots \quad (i)$$

$$\text{In a horizontal direction, the displacement, } x, = vt \quad \dots \quad (ii)$$

Eliminating t between (i) and (ii), we obtain

$$y = \frac{1}{2} \left(\frac{Ee}{m_e} \right) \frac{x^2}{v^2} = \frac{Ee}{2m_e v^2} x^2$$

The path is therefore a *parabola*.

When the electron just passes the plates, $x = D$. The value of y is then $y = EeD^2/2m_e v^2$. The beam then moves in a straight line, as

shown in Fig. 40.8. The time for which the electron is between the plates is D/v . Thus the component of the velocity v_1 , gained in the direction of the field during this time, is given by

$$v_1 = \text{acceleration} \times \text{time} = \frac{Ee}{m_e} \times \frac{D}{v}.$$

Hence the angle θ at which the beam emerges from the field is given by:

$$\tan \theta = \frac{v_1}{v} = \frac{EeD}{m_e v} \cdot \frac{1}{v} = \frac{EeD}{m_e v^2}.$$

The energy of the electron is increased by an amount of $\frac{1}{2}mv_1^2$ as it passes through the plates, since the energy due to the horizontal motion is unaltered.

Deflection in a Magnetic Field

Consider an electron beam, moving with a speed v , which enters a uniform magnetic field of induction B acting perpendicular to the direction of motion. Fig. 40.9. The force F on an electron is then Bev . The direction of the force is perpendicular to both B and v .

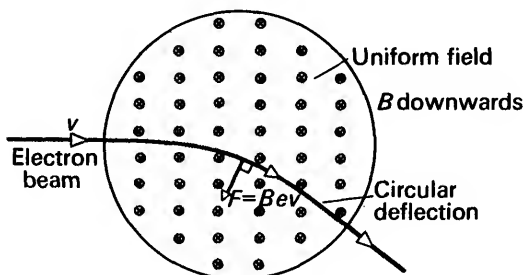


FIG. 40.9. Circular motion in uniform magnetic field.

Consequently, unlike the electric force, the magnetic force cannot change the *energy* of the electron.

The force Bev is always normal to the path of the beam. If the field is uniform, the force is constant in magnitude and the beam then travels in a *circle* of radius r . Since Bev is the centripetal force towards the centre,

$$Bev = \frac{m_e v^2}{r}$$

$$\therefore r = \frac{m_e v}{Be} = \frac{\text{momentum}}{Be}$$

Thomson's Experiment for e/m

In 1897, Sir J. J. Thomson devised an experiment for measuring the ratio *charge/mass* or e/m_e for an electron, sometimes called its *specific charge*.

Thomson's apparatus is shown simplified in Fig. 40.10. C and A are the cathode and anode respectively, and narrow slits are cut in opposite

plates at A so that the cathode rays passing through are limited to a narrow beam. The rays then strike the glass at O, producing a glow there. The rays can be deflected electrostatically by means of connecting a large battery to the horizontal plates P, Q, or magnetically by means of a current passing through Helmholtz coils R, S, on either side of the tube near P and Q, as shown by the small circles in Fig. 40.10.

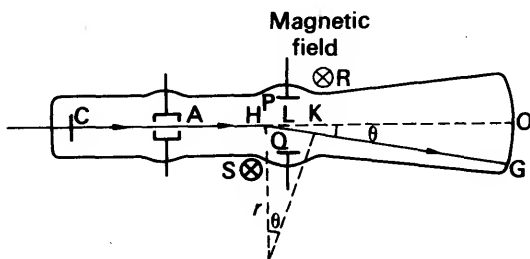


FIG. 40.10. Thomson's determination of e/m for electron (not to scale).

The magnetic field is perpendicular to the paper, and if it is uniform a constant force acts on the cathode rays (electrons) normal to its motion. The particles thus begin to move along the arc HK of a circle of radius r . When they leave the field, the particles move in a straight line and strike the glass at G.

With the usual notation, see p. 1000,

$$\text{force } F = Bev = \frac{m_e v^2}{r},$$

where e is the charge on an electron and m is its mass.

$$\therefore \frac{e}{m_e} = \frac{v}{rB} \quad \dots \quad (i)$$

To find the radius r , we note that, from Fig. 40.10, $\tan \theta = OG/OL = HK/r$.

$$\therefore r = \frac{HK \cdot OL}{OG}.$$

L is about the middle of the solenoid surrounding the plates.

The velocity v was found by applying an electric field between P, Q of such an intensity E as to bring the beam back to O. Then

$$Ee = Bev$$

$$\therefore v = \frac{E}{B}.$$

Thomson found that v was considerably less than the velocity of light, $3 \times 10^8 \text{ m s}^{-1}$, so that cathode rays were certainly not electromagnetic waves.

On substituting for v and r in (i), the ratio charge/mass (e/m_e) for an electron was obtained. Modern determinations show that

$$\frac{e}{m_e} = 1.76 \times 10^{11} \text{ coulomb per kg (C kg}^{-1}\text{)}$$

or
$$\frac{m_e}{e} = \frac{1}{1.76} \times 10^{-11} \text{ kg per coulomb (kg C}^{-1}\text{)} \quad . \quad (ii)$$

Now from electrolysis the electrochemical equivalent of hydrogen is 0.0000104 g per coulomb, or $1.04 \times 10^{-5} \text{ g C}^{-1}$.

$$\therefore \frac{m_H}{e} = 1.04 \times 10^{-8} \text{ kg C}^{-1},$$

assuming the hydrogen ion carries a charge e numerically equal to that on an electron, m_H being the mass of the hydrogen ion. Hence, with (ii),

$$\frac{m_e}{m_H} = \frac{1}{1.76 \times 1.04 \times 10^3} = \frac{1}{1830}.$$

Thus the electron is nearly two thousand times as light as the hydrogen atom.

Until Sir J. J. Thomson's experiment, it was believed that the hydrogen atom was the lightest particle in existence. Note also that in Thomson's experiment, the speed v of the electron beam is measured by means of perpendicular ('crossed') magnetic and electric fields.

Fine Beam Tube

The *fine beam tube* also enables the ratio e/m to be determined. One form of apparatus is shown in Fig. 40.11. It consists of an electron gun which produces an electron beam from a heated cathode. This is accelerated by a voltage V applied to the anode A. The heater and

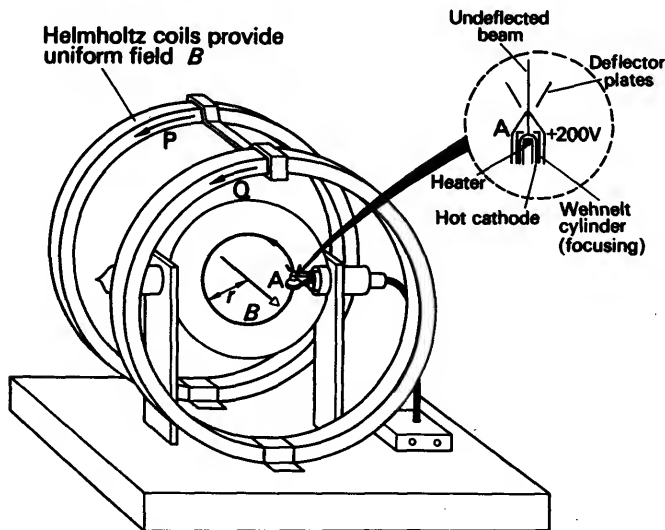


FIG. 40.11. Fine beam tube method for e/m_e .

other voltage supplies are shown in Fig. 40.12. The beam is made visible by having a small quantity of hydrogen gas inside the tube. Collisions of the electrons with the gas molecules cause the latter to emit light, and a straight beam of light is seen, thus showing the electron path.

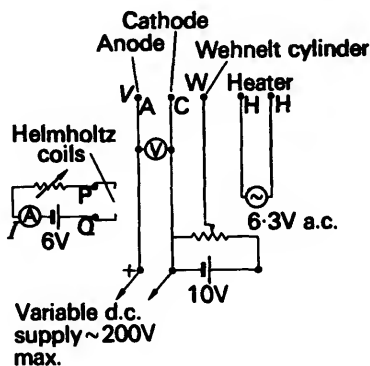


FIG. 40.12. Fine beam tube voltage supplies.

The two large Helmholtz coils P and Q are placed a distance apart equal to their radius to provide a *uniform magnetic field* over the path of the electrons (see p. 936). The value of B can be varied by changing the current I in the coils. Fig. 40.12. When the magnetic field is switched on, the electrons are deflected into a circle of radius r as shown in Fig. 40.11. r may be measured with the aid of a travelling microscope, or by using markers and a mirror. The anode voltage, V , should be measured using a voltmeter. To calculate B , the current I in the Helmholtz coils and their mean radius R are measured. If N is the number of turns in each coil, then from p. 936, B is given by

$$B = 0.72 \frac{\mu_0 NI}{R}$$

Theory

If we assume that the initial velocity of the electrons is zero, the kinetic energy of the electrons in moving through a p.d. V is given by

$$\frac{1}{2} m_e v^2 = eV,$$

where m_e is the mass and v the velocity of each electron.

$$\therefore v^2 = \frac{2eV}{m_e} \quad \dots \quad (i)$$

Since

$$Bev = \frac{m_e v^2}{r},$$

$$\therefore v = \frac{rBe}{m_e} \quad \dots \quad (ii)$$

From (i) and (ii)

$$\frac{2eV}{m_e} = \left(\frac{rBe}{m_e} \right)^2$$

$$\therefore \frac{e}{m_e} = \frac{2V}{r^2 B^2},$$

from which e/m_e may be evaluated if V , r and B are known.

EXAMPLE

Describe and give the theory of a method to determine e the electronic charge. Why is it considered that all electric charges are multiples of e ?

An electron having 450 electron-volts of energy moves at right angles to a uniform magnetic field of magnetic induction (flux density) 1.50×10^{-3} weber metre $^{-2}$. Show that the path of the electron is a circle and find its radius. Assume that the specific charge of the electron is 1.76×10^{11} coulomb kilogramme $^{-1}$. (L.)

With the usual notation, the velocity v of the electron is given by

$$\frac{1}{2}m_e v^2 = eV, \quad \text{where } V \text{ is 450 V.}$$

$$\therefore v = \sqrt{\frac{2eV}{m_e}} \quad \dots \dots \dots (1)$$

The path of the electron is a circle because the force Bev is constant and always normal to the electron path. Its radius r is given by

$$Bev = \frac{m_e v^2}{r}$$

$$\therefore r = \frac{m_e}{e} \cdot \frac{v}{B} = \frac{m_e}{e} \cdot \frac{1}{B} \sqrt{\frac{2eV}{m_e}}, \text{ from (1)}$$

$$\therefore r = \frac{1}{B} \sqrt{2 \frac{m_e}{e} V}$$

Now $e/m_e = 1.76 \times 10^{11}$ C kg $^{-1}$, $V = 450$ V, $B = 1.5 \times 10^{-3}$ T (Wb m $^{-2}$)

$$\therefore r = \frac{1}{1.5 \times 10^{-3}} \sqrt{\frac{2 \times 450}{1.76 \times 10^{11}}} \text{ metre}$$

$$= 4.8 \times 10^{-2} \text{ metre} = 4.8 \text{ cm.}$$

EXERCISES 40

1. Describe an experiment to determine the magnitude of the charge associated with an electron.

State the nature of the path traversed by a charged particle when it is projected at right angles to (a) a uniform magnetic field, (b) a uniform electric field. A uniform magnetic field and a uniform electric field are superimposed so that they allow a charged particle of velocity v to proceed in a straight line in a vacuum. Explain the relations between (i) the directions of the fields and of the particle velocity, (ii) the magnitudes of the fields. (L.)

2. Give an account of a method by which the charge associated with an electron has been measured.

Taking this electronic charge to be -1.60×10^{-19} coulomb, calculate the potential difference in volts necessary to be maintained between two horizontal conducting plates, one 0.50 cm above the other, so that a small oil drop, of mass 1.31×10^{-11} g with two electrons attached to it, remains in equilibrium between them. Which plate would be at the positive potential? (L.)

3. Describe an experiment to determine the ratio of the charge to the mass of electrons. Draw labelled diagrams of (a) the apparatus, (b) any necessary electrical circuits, and show how the result is calculated from the observations.

Two plane metal plates 40 cm long are held horizontally 3.0 cm apart in a vacuum, one being vertically above the other. The upper plate is at a potential of 300 volts and the lower is earthed. Electrons having a velocity of 1.0×10^7 m s $^{-1}$ are injected horizontally midway between the plates and in a direction parallel to the 4.0 cm edge. Calculate the vertical deflection of the electron beam as it emerges from the plates. (e/m for electron = 1.8×10^{11} C kg $^{-1}$.) (N.)

4. Show that if a free electron moves at right angles to a magnetic field the path is a circle. Show also that the electron suffers no force if it moves parallel to the field. Point out how the steps in your proof are related to fundamental definitions.

If the path of the electron is a circle, prove that the time for a complete revolution is independent of the speed of the electron.

In the ionosphere electrons execute 1.4×10^6 revolutions in a second. Find the strength of the magnetic induction B in this region. (Mass of an electron = 9.1×10^{-28} g; electronic charge = 1.6×10^{-19} coulomb.) (C.)

5. The electron is stated to have a mass of approximately 10^{-27} g and a negative charge of approximately 1.6×10^{-19} C. Outline the experimental evidence for this statement. Formulae may be quoted without proof. You are not required to justify the actual numerical values quoted.

An oil drop of mass 3.25×10^{-12} g falls vertically, with uniform velocity, through the air between vertical parallel plates which are 2 cm apart. When a p.d. of 1000 V is applied to the plates the drop moves towards the negatively charged plate, its path being inclined at 45° to the vertical. Explain why the vertical component of its velocity remains unchanged and calculate the charge on the drop.

If the path of the drop suddenly changes to one at $26^\circ 30'$ to the vertical, and subsequently to one at 37° to the vertical, what conclusions can be drawn? (O. & C.)

6. Give an account of Millikan's experiment for determining the value of the electronic charge e .

In a Millikan-type apparatus the horizontal plates are 1.5 cm apart. With the

electric field switched off an oil drop is observed to fall with the steady velocity $2.5 \times 10^{-2} \text{ cm s}^{-1}$. When the field is switched on the upper plate being positive, the drop just remains stationary when the p.d. between the two plates is 1500 volts.

(a) Calculate the radius of the drop. (b) How many electronic charges does it carry? (c) If the p.d. between the two plates remains unchanged, with what velocity will the drop move when it has collected two more electrons as a result of exposure to ionizing radiation? (O. & C.)

7. Give a short account of the phenomena observed when an electric discharge passes through a gas at very low pressure (10^{-6} atmosphere). Describe very briefly experiments which reveal the nature of the discharge.

What is the direction of the force acting on a negatively charged particle moving through a magnetic field? Deduce the shape of the path of a charged particle projected at right angles to a uniform magnetic field. (L.)

8. Describe a method for measuring the charge per unit mass for the electron, showing how the value is calculated from the observations.

An ion, for which the charge per unit mass is $4.40 \times 10^7 \text{ C kg}^{-1}$, has a velocity of $3.52 \times 10^7 \text{ cm s}^{-1}$ and moves in a circular orbit in a magnetic field of induction 0.4 Wb m^{-2} . What will be the radius of this orbit? (L.)

9. Describe an oil drop method of determining the electronic charge, e . How may Avogadro's constant be found when e is known?

An oil drop of radius $1.000 \times 10^{-3} \text{ cm}$ falls freely in air, midway between two vertical parallel metal plates of large extent, which are 0.5000 cm apart, and its terminal velocity is 1.066 cm s^{-1} . When a potential difference of 3000 volts is applied between the plates, the path of the drop becomes a straight line inclined at an angle of $31^\circ 36'$ to the vertical. Find the charge on the drop. (Assume the viscosity of air to be $1.816 \times 10^{-5} \text{ kg m}^{-1} \text{ s}^{-1}$.) (L.)

10. An electron with a velocity of 10^7 m s^{-1} enters a region of uniform magnetic flux density of 0.10 Wb m^{-2} , the angle between the direction of the field and the initial path of the electron being 25° . By resolving the velocity of the electron find the axial distance between two turns of the helical path. Assume that the motion occurs in a vacuum and illustrate the path with a diagram. ($e/m = 1.8 \times 10^{11} \text{ coulomb kg}^{-1}$.) (N.)

11. Describe an apparatus for determining the ratio of the charge e to the mass m of the electron. Explain how measurements are carried out with the apparatus, and derive the relationship between e/m and the experimentally measured quantities.

Indicate briefly how you would attempt to test whether the particles emitted in the photoelectric effect and in thermionic emission are the same.

When low energy electrons are moving at right angles to a uniform magnetic field of flux density $10^{-3} \text{ Wb m}^{-2}$, they describe circular orbits 2.82×10^7 times per second. Deduce a value for e/m . (O. & C.)

12. Describe and give the theory of the Millikan oil drop experiment for the determination of the electronic charge. What is the importance of the experiment?

In one such experiment a single charged drop was found to fall under gravity at a terminal velocity of $0.0040 \text{ cm per second}$ and to rise at $0.0120 \text{ cm per second}$ when a field of $2 \times 10^5 \text{ volt per m}$ was suitably applied. Calculate the electronic charge given that the radius, a , of the drop was $6.0 \times 10^{-7} \text{ m}$ and that the viscosity, η , of the gas under the conditions of the experiment was $180 \times 10^{-5} \text{ N s m}^{-2}$ (N.)

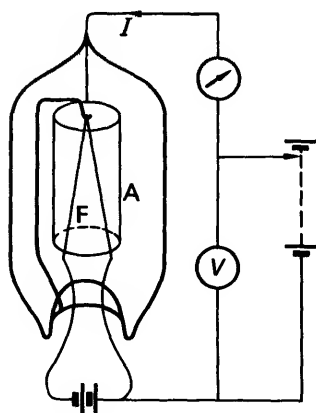
chapter forty-one

Radio Valves. C.R.O. Junction Diode. Transistor

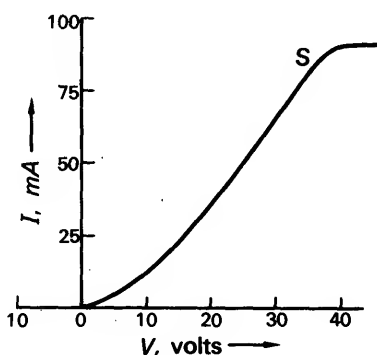
RADIO VALVES. CATHODE RAY OSCILLOGRAPH

Rectification by A.C. Diode Valve

ALTERNATING current is easier to distribute than direct current, because alternating voltage can be transformed easily up or down. For electrolysis, battery-charging, and the operation of radio-receivers and transmitters, however, direct current is essential. It can be obtained from an alternating current supply by means of a *rectifier*, which is a device that will only pass current in one direction. A common type of rectifier is that called a *diode valve*. It contains a metal filament, F in Fig. 41.1 (a), surrounded by a metal anode A. The filament is heated by a current drawn from a low voltage supply, and emits electrons. A circuit for varying the anode potential is shown in Fig. 41.1 (a).



(i) Diode



(ii) Characteristic

FIG. 41.1. Diode valve.

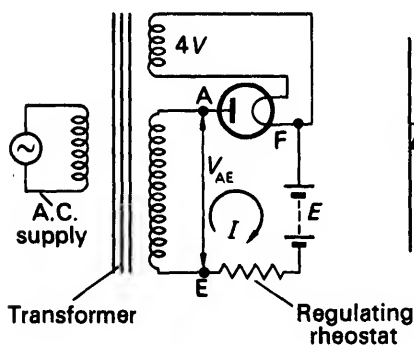
Since electrons are negative charges, such a device passes current when its anode is made positive with respect to its filament, but not when the anode is made negative. Fig. 41.1 (b) shows the curve of anode current against anode potential for a small diode; it is called the diode's characteristic curve, or simply its *characteristic*. The current increases with the positive anode potential as far as the point S. Beyond this

point the current does not increase, because the anode is collecting all the electrons emitted by the filament; the current is said to be saturated.

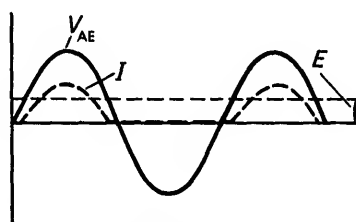
At first sight we might expect that any positive anode potential, however small, would draw the full saturation current from the filament to the anode. But it does not, because the charges on the electrons make them repel one another. Thus the cloud of electrons between the anode and filament repels the electrons leaving the filament, and turns some of them back. The electrons round the filament are like the molecules in a cloud of vapour above a liquid; they are continually escaping from it and returning to it. The positive anode draws some away from the cloud, as a wind carries water vapour away from a pool. The wind, or the anode, thins out the cloud, so that more electrons or molecules escape than return. The higher the anode potential, the fewer electrons return to the filament; as the anode potential rises, the current increases, to its saturation limit.

Rectifier Circuit

When a diode is used as a rectifier it is connected in a circuit such as Fig. 41.2 (i). The low-voltage secondary of the transformer simply provides the heating current from the filament. The current to be rectified is drawn from the high-voltage secondary. One end of this



(i) Circuit



(ii) Voltages and current

FIG. 41.2. Rectifier circuit with diode.

secondary is connected to the load, which could be, as shown, an accumulator on charge; the other end of the load is connected to the anode of the diode, and the other end of the secondary to one of the filament connexions. When the transformer secondary voltage V_{AE} is greater than the e.m.f. E of the accumulator, the anode is positive with respect to the filament; electrons from the hot filament are then drawn to the anode, and a current flows through the load (Fig. 41.2 (ii)). On half-cycles when the anode is negative, the electrons are repelled, and no current flows. Because it only allows current to flow through it in one direction, a thermionic diode is often called a valve.

Some rectifying valves contain a little mercury vapour. When electrons flow through them, they ionize the mercury atoms, as explained on p. 1016. The ions and electrons thus produced make the valve a very good conductor, and reduce the voltage drop across it; they therefore allow more of the voltage from the transformer to appear across the load.

The current from a rectifier flows in pulses, whenever the anode is positive with respect to the filament. Sometimes a smoother current is required, as, for example, in a radio-receiver, where the pulses would cause a humming sound in the loudspeaker. The current can be smoothed by connecting an inductance coil of about 30 henrys in series with the load. The inductance prevents rapid fluctuations in current. So also does a capacitor of about 16 microfarads connected across the load. Generally the two are used together to give a very smooth output.

Metal Rectifier

Yet other rectifiers are not thermionic at all. One such type consists of an oxidized copper disc, $\text{Cu}_2\text{O}/\text{Cu}$, pressed against a disc of lead, Pb (Fig. 41.3 (i)). These conduct well when the lead is made positive, but very badly when it is made negative; they are called metal rectifiers. A

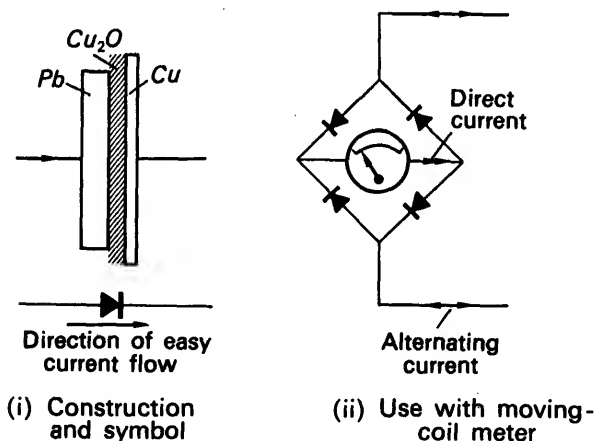


FIG. 41.3. Metal rectifier and use.

metal rectifier can be used to convert a moving-coil milliammeter into an alternating-current ammeter or voltmeter (Fig. 41.3 (ii)). Such a meter is more sensitive than a moving-iron or hot-wire instrument, and has a more open scale near zero: its deflection is roughly proportional to the average value of the current or voltage.

Cathode-Ray Oscillograph

An oscillograph is an instrument for plotting one varying physical quantity—potential difference, sound-pressure, heart-beat—against

another—current, displacement, time. A cathode-ray oscillograph, of the kind we are about to describe, plots alternating potential difference against time. It is so called because it traces the desired wave-form with a beam of electrons, and beams of electrons were originally called cathode rays.

A cathode-ray oscillograph is essentially an electrostatic instrument. It consists of a highly evacuated glass tube, T in Fig. 41.4, one end of

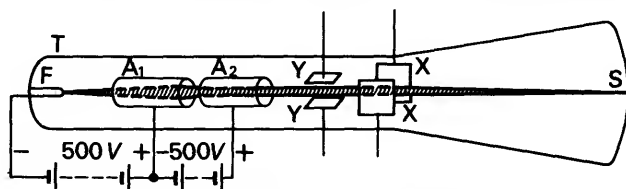


FIG. 41.4. A cathode-ray oscillograph tube.

which opens out to form a screen S which is internally coated with zinc sulphide. A hot filament F, at the other end of the tube, emits electrons. These are then attracted by the cylinders A_1 and A_2 , which have increasing positive potentials with respect to the filament. Many of the electrons, however, shoot through the cylinders and strike the screen; where they do so, the zinc sulphide fluoresces in a green spot. On their way to the screen, the electrons pass through two pairs of metal plates, XX and YY, called the deflecting plates.

Deflection; Time-base

If a battery were connected between the Y-plates, so as to make the upper one positive, the electrons in the beam would be attracted towards that plate, and the beam would be deflected upwards. In the same way, the beam can be deflected horizontally by a potential difference applied between the X-plates. When the oscillograph is in use, the alternating potential difference to be examined is applied between the Y-plates. If that were all, then the spot would be simply drawn out into a vertical line. To trace the wave-form of the alternating potential difference, the X-plates are used to provide a time-axis. A special valve circuit generates a potential difference which rises steadily to a certain value, as shown in Fig. 41.5 (i), and then falls rapidly

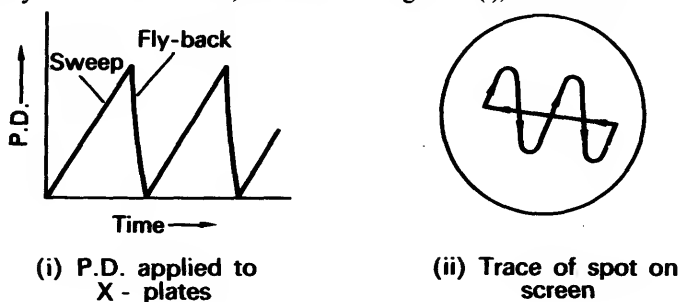


FIG. 41.5. Action of a C.R.O.

to zero; it can be made to go through these changes tens, hundreds, or thousands of times per second. This potential difference is applied between the X-plates, so that the spot is swept steadily to the right, and then flies swiftly back and starts out again. This horizontal motion provides what is called the time-base of the oscillograph. On it is superimposed the vertical motion produced by the Y-plates; thus, as shown in Fig. 41.5 (ii), the wave-form of the potential difference to be examined is displayed on the screen.

Focusing

To give a clear trace on the screen, the electron beam must be focused to a sharp spot. This is the function of the cylinders A_1 and A_2 , called the first and second anodes. Fig. 41.6 shows the equipotentials of the field between them, when their difference of potential is 500 volts.

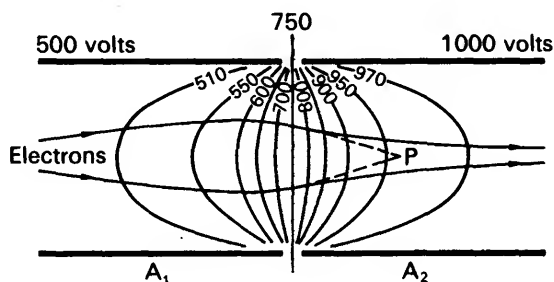


FIG. 41.6. Focusing in a C.R.O. tube.

Electrons entering the field from the filament experience forces from low potential to high at right angles to the equipotentials. They have, however, considerable momentum, because they have been accelerated by a potential difference of about 500 volts, and are travelling fast. Consequently the field merely deflects them, and, because of its cylindrical symmetry, it converges the beam towards the point P. Before they can reach this point, however, they enter the second cylinder. Here the potential rises from the axis, and the electrons are deflected outwards. However, they are now travelling faster than when they were in the first cylinder, because the potential is everywhere higher. Consequently their momentum is greater, and they are less deflected than before. The second cylinder, therefore, diverges the beam less than the first cylinder converged it, and the beam emerges from the second anode still somewhat convergent. By adjusting the potential of the first anode, the beam can be focused upon the screen, to give a spot a millimetre or less in diameter.

Electron-focusing devices are called electron-lenses, or electron-optical systems. For example, the action of the anodes A_1 and A_2 is roughly analogous to that of a pair of glass lenses on a beam of light, the first glass lens being converging, and the second diverging, but weaker.

Uses of Oscillograph

In addition to displaying waveforms, the oscillograph can be used for measurement of voltage, frequency and phase.

1. A.C. voltage

An unknown a.c. voltage, whose peak value is required, is connected to the Y-plates. With the time-base switched off, the vertical line on the screen is centred and its length then measured. Fig. 41.7 (i). This is proportional to twice the amplitude or peak voltage, V_0 . By measuring the length corresponding to a known a.c. voltage V , then V_0 can be found by proportion.

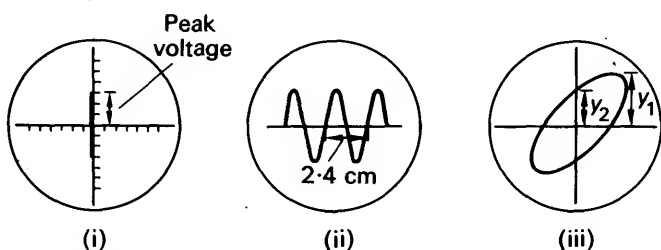


FIG. 41.7. Uses of oscillograph.

Alternatively, using the same gain, the waveforms of the unknown and known voltages, V_0 and V , can be displayed on the screen. The ratio V_0/V is then obtained from measurement of the respective peak-to-peak heights.

2. Comparison of frequency

If a calibrated time-base is available, frequency measurements can be made. In Fig. 41.7 (ii), for example, the trace shown is that of an alternating waveform with the time-base switched to the '5 millisecc/cm' scale. This means that the time taken for the spot to move 1 cm horizontally across the screen is 5 milliseconds. The horizontal distance on the screen for one cycle is 2.4 cm. This corresponds to a time of 5×2.4 ms or 12.0 ms = 12×10^{-3} seconds, which is the period T .

$$\therefore \text{frequency} = \frac{1}{T} = \frac{1}{12 \times 10^{-3}} = 83 \text{ Hz.}$$

If a comparison of frequencies f_1, f_2 is required, then the corresponding horizontal distances on the screen are measured. Suppose these are d_1, d_2 respectively. Then, since $f \propto 1/T$,

$$\frac{f_1}{f_2} = \frac{T_2}{T_1} = \frac{d_2}{d_1}.$$

3. Measurement of phase

The use of a double beam oscillograph to measure phase difference is given on p. 586. If only a single beam tube is available, an elliptical trace can be obtained. With the time-base switched off, one input is joined to the X-plates and the other to the Y-plates. We consider

only the case when the frequencies of the two signals are the same. An ellipse will then be seen generally on the screen, as shown in Fig. 41.7 (iii) (see p. 1014).

The trace is centred, and the peak vertical displacement y_2 at the middle O, and the peak vertical displacement y_1 of the ellipse, are then both measured. Suppose the x -displacement is given by $x = a \sin \omega t$, where a is the amplitude in the x -direction, and the y -displacement by $y = y_1 \sin (\omega t + \phi)$, where y_1 is the amplitude in the y -direction and ϕ is the phase angle. When $x = 0$, $\sin \omega t = 0$, so that $\omega t = 0$. In this case, $y = y_2 = y_1 \sin \phi$. Hence $\sin \phi = y_2/y_1$, from which ϕ can be found.

TRIODE VALVE—AMPLIFIER, OSCILLATOR, DETECTOR

Triode Valve

A few years after the invention of the diode valve Lee de Forest introduced the triode valve. This had three electrodes: a cathode C, the emitter of electrons; an anode A, the collector of electrons; and a *grid* G, a wire with open spaces, placed between the anode and cathode (Fig. 41.8 (i)). The function of the grid is to control the electron flow to the anode, and for this purpose the grid has a small negative potential relative to the cathode. The grid is nearer the cathode than the anode, and its potential thus affects the electric field round the cathode more, with the result that the grid potential has a more delicate control than the anode potential over the anode current. As we shall see later, this enables the triode to act as an amplifier of alternating voltages as well as a detector.

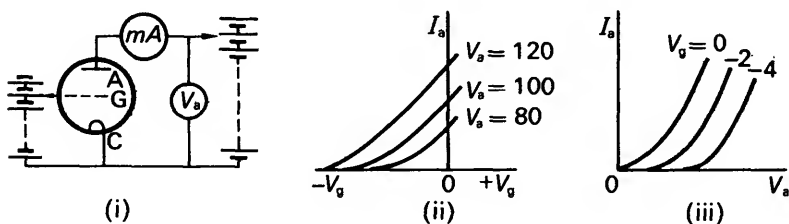


FIG. 41.8. Triode valve characteristics.

Triode Valve Characteristics

In order to predict the performance of a valve in a circuit, the 'characteristics' of the valve must be first determined. The chief characteristics are I_a-V_g (V_a constant), the variation of anode current with grid voltage when the anode voltage is constant; and I_a-V_a (V_g constant), the variation of anode current with anode voltage when grid voltage is constant. The I_a-V_g curves are known as the valve's *mutual characteristics*; the I_a-V_a curves as the *anode characteristics*.

The mutual characteristics obtained are shown in Fig. 41.8 (ii). When

the anode voltage is 80 volts, a negative voltage on the grid such as -15 volts creates a resultant negative electric intensity at the cathode, and hence no electrons flow past the grid. As the negative voltage is reduced and reaches a certain value the attractive effect of the positive anode voltage overcomes the repulsive effect of the grid voltage, and electrons now reach the anode. As the negative voltage is reduced further, more electrons reach the anode, and the current increases as shown. The general shape of the I_a - V_g curves is an initial curvature, followed by a straight line.

The anode characteristics, I_a - V_a curves, are shown in Fig. 41.8 (iii), and are explained in a similar way. As the anode voltage, V_a , increases, the anode current increases. Generally, the anode current begins to flow at higher values of V_a when the grid voltage is increased more negatively. If the anode voltage is increased sufficiently, all the electrons emitted by the cathode are collected, and the current has then reached its saturation value (p. 1010).

Valve Constants

There are three main constants or properties of a radio valve. These are:

1. *Anode or A.C. Resistance*, R_a , which is defined by

$$R_a = \frac{\delta V_a}{\delta I_a} (V_g \text{ constant}),$$

the changes in V_a and I_a being taken on the *straight* part of the anode characteristics.

2. *Mutual conductance*, g_m , which is defined by

$$g_m = \frac{\delta I_a}{\delta V_g} (V_a \text{ constant}),$$

the changes in I_a and V_g being taken on the straight part of the mutual characteristics.

3. *Amplification factor*, μ , which is defined by

$$\mu = \frac{\delta V_a}{\delta V_g},$$

where δV_a produces the same change in anode current (V_g constant) as δV_g (V_a constant).

Thus, generally, R_a is the 'resistance' of the valve when the anode circuit variations are considered, g_m is the change in anode current produced by unit grid voltage variation, and μ is a measure of the 'step-up' in voltage produced in the anode circuit by a change in the grid voltage.

Triode as Voltage Amplifier

When a valve is used as a voltage amplifier in radio circuits, it is important to realize at the outset that it amplifies *alternating* voltages, and that these voltages are applied in the grid-cathode circuit, as

represented by V in Fig. 41.9 (i). The action of the valve should not only result in an increased alternating voltage V_0 in the anode circuit, known as the 'output voltage', but the waveform of V_0 should be exactly the same as V , the applied voltage, so that there is no distortion. In order to obtain no distortion, a steady negative p.d. (grid-bias, G.B.) is also connected in the grid-cathode circuit, as shown in Fig. 41.9 (i).

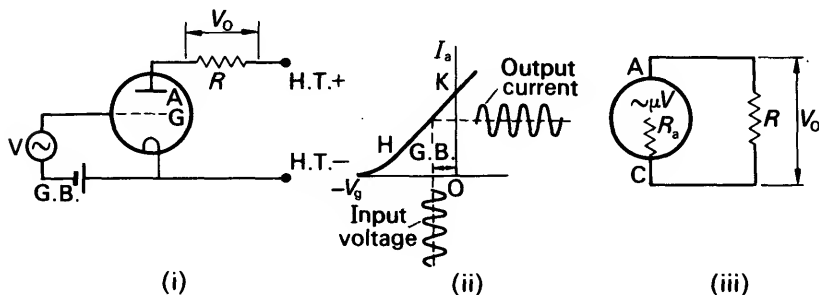


FIG. 41.9. Triode amplification.

The most suitable value of the grid-bias p.d. is OX volts, where X (not shown) corresponds to the middle of the straight part HK of the I_a-V_g characteristic (Fig. 41.9 (ii)). Then, if the applied alternating voltage V has a peak value less than OX, the actual grid potential values will produce anode current variations corresponding to the straight part of the characteristic. The anode or output current will then have a waveform exactly the same as the applied or input voltage V (Fig. 41.9 (ii)). As we shall now show, the triode acts as a 'voltage amplifier' in this case.

Voltage Gain or Amplification Factor

The magnitude of the voltage amplification can be found by replacing the valve circuit in Fig. 41.9 (i) by an 'equivalent A.C. circuit'. Since a change of p.d. δV_g in the grid-cathode circuit is equivalent to a change of $\mu\delta V_g$ in the anode circuit, the alternating voltage V is equivalent to an alternating voltage μV in the anode circuit. We therefore consider that, between the anode and cathode, the valve is an alternating voltage generator of e.m.f. μV , with a internal resistance R_a , the a.c. resistance discussed on p. 1016. See Fig. 41.9 (iii).

To convert the alternating current in the anode circuit to an alternating voltage, a large resistance R is needed, of the order of thousands of ohms. The internal resistance of the H.T. battery and that of the G.B. battery can be neglected by comparison, and since varying voltages are now considered, the magnitudes of the steady H.T. and G.B. voltages can also be ignored. The complete valve equivalent a.c. circuit is therefore as shown in Fig. 41.9 (iii).

The total resistance of the circuit is $R + R_a$. Thus the alternating current, I , flowing

$$= \frac{\mu V}{R + R_a}.$$

$$\therefore \text{output alternating voltage, } V_0 = IR = \frac{\mu VR}{R + R_a}$$

$$\therefore \text{voltage gain or amplification factor} = \frac{V_0}{V} = \frac{\mu R}{R + R_a} \quad (i)$$

Thus a triode has an amplification factor μ of 10, an internal resistance R_a of 8000 ohms and a resistance R of 10000 ohms,

$$\text{voltage gain} = \frac{\mu R}{R + R_a} = \frac{10 \times 10000}{10000 + 8000} = 5.6.$$

Hence if the applied alternating voltage is 0.2 volt (r.m.s.),

$$\text{amplified output voltage} = 5.6 \times 0.2 = 1.1 \text{ volts (r.m.s.).}$$

Couplings

Fig. 41.10 shows how the output voltage V_0 across R is passed from the valve V1 to the next valve V2. Fig. 41.10 (i) illustrates *resistance-capacitance* coupling. Here one end of the load R is joined to the coupling capacitor, C_g . The other end of R is joined to the lower end of the resistor R_g through the relatively low internal resistance of the h.t. battery. Thus, from an a.c. point of view, C_g and R_g are effectively in

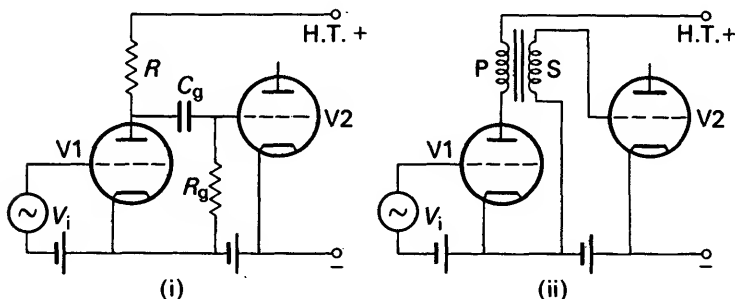


FIG. 41.10. Couplings: (i) Resistance-capacitance; (ii) Transformer.

parallel with R . R_g thus passes to V2 a fraction of the output (a.c.) voltage across R . Note that C_g is necessary to isolate the grid of V2 from the positive potential of the h.t. battery.

Fig. 4.10 (ii) illustrates *transformer coupling*. The load in the anode circuit is the primary coil P. The output (a.c.) voltage across this coil is amplified by the transformer and passed to the next valve V2 by the secondary coil S.

Basic Oscillatory Circuit

About 1862 Lord Kelvin showed theoretically that when an electrical disturbance is made in a capacitor consisting of a capacitor and a coil and then left, *oscillations* of current occur (see also p. 578). Thus suppose a current I flows at an instant t in a circuit consisting of a coil of inductance L and negligible resistance, in series with a capacitor of

capacitance C (Fig. 41.11). Then, from p. 924, if Q is the charge on the capacitor,

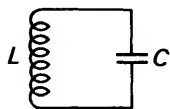


FIG. 41.11.
Basic oscillatory
circuit.

$$\text{p.d. across inductance} = -L \frac{dI}{dt}$$

$$= \text{p.d. across capacitor} = \frac{Q}{C}$$

$$\therefore -L \frac{dI}{dt} = \frac{Q}{C}$$

$$\text{But } I = \frac{dQ}{dt} \therefore -L \frac{d^2Q}{dt^2} = \frac{Q}{C}$$

$$\therefore \frac{d^2Q}{dt^2} = -\frac{1}{LC} \cdot Q \quad (\text{i})$$

This is a 'simple harmonic' equation. Thus Q , the charge circulating, varies with time t according to the relation

$$Q = Q_0 \sin \omega t, \quad (\text{ii})$$

where Q is the maximum value of the varying charge and ω is a constant given by $\omega^2 = 1/LC$, or

$$\omega = \frac{1}{\sqrt{LC}}$$

The frequency, f , of the oscillatory charge is given by

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi\sqrt{LC}} \quad (\text{iii})$$

A coil-capacitor series circuit is thus a basic oscillatory circuit, and the frequency of the oscillations of charge (or current) depends on the magnitudes of the inductance L and capacitance C .

The physical reason for the oscillations is the constant interchange of energy between the capacitor and the coil. When the bob of a pendulum is oscillating, its energy at the end of a swing is wholly potential. This gradually changes into kinetic energy until it is wholly kinetic at the middle of the swing, and then becomes potential energy again at the end of the swing. In a similar way, the capacitor becomes fully charged at one instant, the energy being electrostatic energy, and as the capacitor discharges, the energy is stored in the magnetic field of the coil. When the capacitor is fully discharged, the energy is wholly in the magnetic field. After this, the capacitor charges up the other way round, storing electrical energy, and when it is fully charged, there is then no energy in the coil's magnetic field. See p. 578.

Damped and Undamped Waves

The oscillations of charge or current in a circuit containing only an inductor and a capacitor will theoretically last indefinitely with a constant amplitude (Fig. 41.12 (i)). In practice, however, some of the energy is dissipated as heat in the resistance of the coil. Since this

energy is no longer available as oscillatory energy, the amplitude of the oscillations gradually diminish, and a *damped* oscillation is thus obtained (Fig. 41.12 (ii)). This is analogous to the case of a vibrating

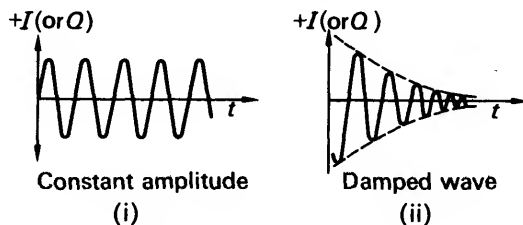


FIG. 41.12. Undamped (continuous) waves and damped waves.

tuning-fork. If there were no air friction the oscillations of the prongs would proceed indefinitely with constant amplitude, but in practice the amplitude of the oscillations gradually diminishes to zero.

Triode as an Oscillator

A triode valve can be used to maintain oscillations in a coil (L , R)-capacitor (C) electrical circuit. The principle of the method is shown in Fig. 41.13, which consists of the basic oscillatory circuit connected in the grid-cathode circuit of the valve; a coil L_1 , close to the oscillator coil, in the anode circuit; a high-tension (h.t.) battery; and a capacitor C_g and a resistance R_g in the grid circuit, for a reason to be explained later.

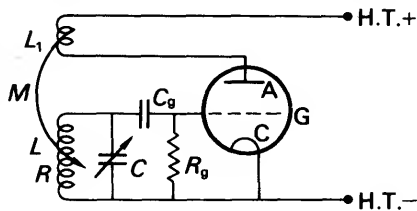


FIG. 41.13. Triode as oscillator.

When the circuit is made, oscillations of current occur in the coil (L , R)-capacitor (C) circuit, as already explained. The oscillatory coil alternating p.d., across C , is amplified by the valve, and oscillatory currents are then obtained in the anode circuit and hence in the coil L_1 . By mutual induction M between the coils, *some energy is fed back to the oscillator circuit*. If the feed-back is correctly phased and is of the required amount, it will help to maintain the oscillations of current in the oscillatory current, which will then become undamped. The magnitude and phase of the feed-back can be varied by altering the position of the coil L_1 and, if necessary, reversing its connections in the anode circuit. With audio-frequency oscillations, a continuous whistle can usually be heard.

Efficiency of Oscillator

The source of the oscillatory energy is the h.t. battery in the circuit. An oscillator is thus said to be a device for converting d.c. energy into a.c. energy. The *efficiency* of the circuit is defined as:

$$\frac{\text{output (oscillatory) energy, a.c.}}{\text{input energy, d.c.}} \times 100\%$$

Once the balance-wheel of a watch is set in motion, energy is imparted to it regularly only at certain times of its oscillation. In this way the balance-wheel is supplied with the least amount of energy needed to maintain undamped oscillations. For the same reason, the triode oscillator circuit has a capacitor C_g and a high-resistance R_g in the grid-circuit. See Fig. 41.13. When the circuit is first made the p.d. across R_g is zero, and hence the grid is at zero potential. The alternating or oscillatory voltage across the capacitor C makes the grid positive in potential for some part of its cycle. Some electrons are therefore drawn into the grid circuit, charging the capacitor C_g . During the oscillatory voltage some charge (electrons) leaks away through R_g and

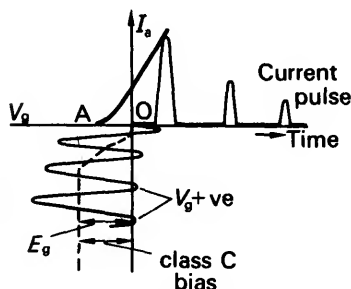


FIG. 41.14. Variation of grid p.d. and anode current (*not to scale*).

the grid thus becomes more negative in potential. Fig. 41.14 shows roughly how the grid potential decreases, and with suitably chosen values of C_g and R_g it soon settles down to some steady negative value E_g , which is the grid-bias while the valve is functioning. In this condition the oscillatory voltage across the capacitor C only produces a pulse of current in the anode circuit at brief intervals, as shown in Fig. 41.14, and by mutual induction, energy is fed

back simultaneously by the coil L_1 into the oscillatory circuit (L , R and C) to make up for the energy lost as heat in the resistance R .

The valve is here said to be operating under 'class C' conditions, that is, the negative grid-bias E_g is at least twice the grid-bias value OA which cuts off the anode current in the I_a - V_g characteristic (Fig. 41.14). During part of the cycle the grid potential becomes positive, as shown. A fixed negative grid-bias equal to E_g is unsuitable in a valve oscillator circuit. The alternating (oscillatory) voltage across the capacitor C , once obtained, would not produce any current in the anode circuit and oscillations would then not continue owing to lack of feed-back of energy.

Radio Waves and Modulation

In 1887 Hertz found by experiment that when an oscillatory voltage of high frequency was connected to two capacitor plates far apart, some of the oscillatory energy travelled in space some distance from the plates and was detected. This was the first discovery of the existence of *radio waves*. A transmitting aerial is a form of capacitor in which one 'plate' is high above the other 'plate', the earth. Theory and experiment show that radio waves will not travel out far from a transmitting aerial unless their frequency is very high. Valve oscillator circuits (p. 1020) therefore usually produce alternating voltages of the order of a million (10^6) Hz, 1 MHz, or more, known as *radio-frequencies (R.F.)*. The Radio 4 station in Britain, broadcasting on 330 metres wavelength, sends out radio waves of a frequency of 908000 Hz (908 kHz). Another

station sends out radio waves of very high frequency (V.H.F.) 90.0 MHz.

At broadcasting stations, the oscillator alone would produce a radio wave of constant amplitude (Fig. 41.15 (i), (ii)). When *audio-frequency*

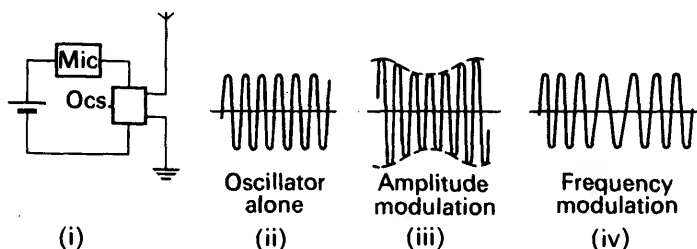


FIG. 41.15. Amplitude and frequency modulation.

(A.F.) currents, due to speech or music, are fed through a microphone into the oscillatory circuit, the radio waves are affected or 'modulated' accordingly. In *amplitude modulation* (A.M.), the amplitude of the radio-frequency wave varies exactly as the audio-frequency (Fig. 41.15 (iii)). In *frequency modulation* (F.M.) the amplitude of the radio-frequency wave is constant but the audio-frequency is superimposed on the frequency of the radio wave (Fig. 41.15 (iv)).

Diode Valve Detection

The principle of the diode valve was discussed on p. 1009. There we showed that a diode valve, which consists of a nickel plate or anode placed in a vacuum opposite a cathode emitting electrons, allowed current to flow through it only when the anode was positive in potential relative to the cathode.

The diode can be used to convert alternating to direct voltage (see p. 1010). It can also be used to 'detect' the audio-frequency variation carried along with the modulated wave sent out by transmitters. If a modulated wave is applied between the anode and cathode of a diode, with a resistor R in the circuit (Fig. 41.16 (i)), the valve conducts on the

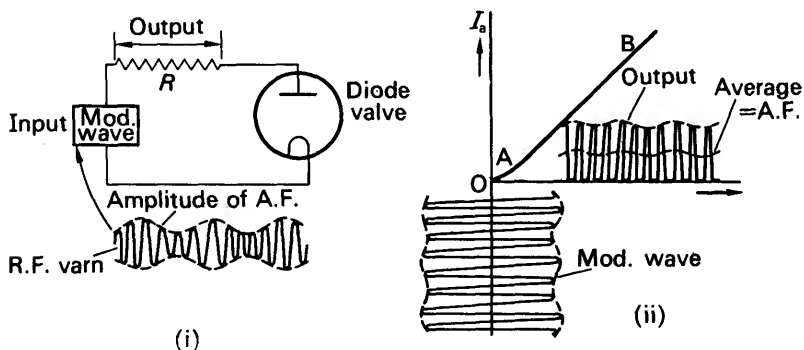


FIG. 41.16. Diode valve detection.

positive parts of the cycle. The variation of current I_a in the anode circuit, the output current, is then as shown in Fig. 41.16 (ii), where OAB is the I_a - V_a curve. The average value of the current, it will be noted, follows the variation of the amplitude of the modulated wave, and hence the voltage across R , called the *output voltage*, has the same audio-frequency variation. In this way the diode is said to act as a 'detector' of the audio-frequency. If the modulated wave were applied to the resistance R without using the diode, the average current obtained would be zero.

Triode as a Detector

We have just shown that the diode can act as a 'detector' of the audio-frequency carried with a modulated wave. The triode can also act as a detector, and in one method, known as *anode-bend detection*, the modulated wave is applied in the grid-cathode circuit, together with a steady grid-bias (G.B.) corresponding to a point on the bend of the I_a - V_g characteristic (Fig. 41.17 (i), (ii)). The swings of the modu-

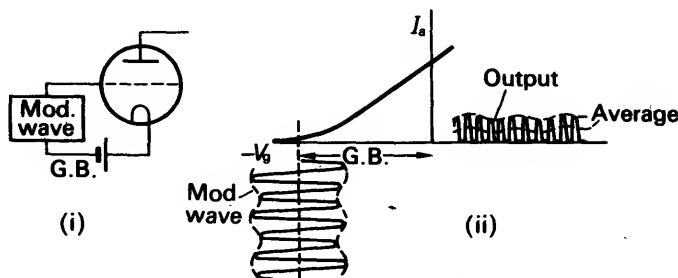


FIG. 41.17. Triode detection.

lated wave on one half of the cycles now produce an anode-current variation, but very little current flows on the other half-cycles. The output, or anode current, thus varies as shown in Fig. 41.17 (ii), and hence the *average* current variation follows the variation of the peaks of the current, which is the audio-frequency variation carried along by the modulated wave (see p. 1022). By means of high resistance earphones and a suitable capacitor across it, the audio-frequency variation can be heard.

JUNCTION DIODE. TRANSISTOR

Semiconductors

In receivers, the radio valve has been superseded by components made from *semiconductors*, which perform the same function as the valve. Semiconductors are a class of solids with electrical resistivity between that of a conductor and an insulator. For example, the resistivity of a conductor is of the order $10^{-8} \Omega \text{ m}$, that of an insulator is $10^4 \Omega \text{ m}$ and higher, and that of a semiconductor is $10^{-1} \Omega \text{ m}$. Silicon

and germanium are examples of semiconductor elements widely used in industry.

Electrons and Holes

Silicon and germanium atoms are tetravalent. They have four electrons in their outermost shell, called *valence electrons*. One valence electron is shared with each of four surrounding atoms in a tetrahedral arrangement, forming 'covalent bonds' which maintain the crystalline solid structure. Fig. 41.18 (i) is a two-dimensional representation of the structure.

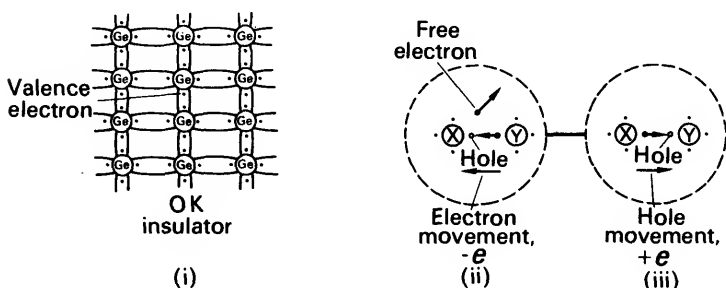


FIG. 41.18. Semiconductor. Electron ($-$) and hole ($+$) movement.

At 0K, all the valence electrons are firmly bound to the nucleus of their particular atom. At room temperature, however, the thermal energy of a valence electron may become greater than the energy binding it to its nucleus. The covalent bond is then broken. The electron leaves the atom, X say, and becomes a free electron. This leaves X with a vacancy or *hole*. Fig. 41.18 (ii). Since X now has a net positive charge, an electron in a neighbouring atom may then be attracted. Thus the hole appears to move to Y.

The hole movement through a semiconductor is random. But if a battery is connected, the valence electrons are urged to move in one direction and to fill the holes. The holes then drift in the direction of the field. Thus the holes move as if they were carriers with a positive charge $+e$, where e is the numerical value of the charge on an electron. Fig. 41.18 (iii). The current in the semiconductor is also carried by the free electrons present. These are equal in number to the holes in a pure semiconductor and drift in the opposite direction since they are negative charges. The mobility of an electron, its average velocity per unit electric field intensity, is usually much greater than that of a hole.

In electrolytes (p. 844), the current is also carried by moving negative and positive charges but the carriers here are *ions*. It should be noted that, in a pure semiconductor, there are equal numbers of electrons and holes. *Electron-hole pairs* are said to be produced by the movement of an electron from bound state in an atom to a higher energy level, where it becomes a free electron.

Effect of Temperature Rise

In contrast to a semiconductor, the carriers of electricity in a metal such as copper are only free electrons. Further, as the temperature of the metal rises, the amplitude of vibration of the atoms increases and more 'collisions' with atoms are then made by drifting electrons. Thus, as stated on p. 837, the resistance of a pure metal increases with temperature rise.

In the case of a semiconductor, however, the increase in thermal energy of the valence electrons due to temperature rise enables more of them to break the covalent bonds and become free electrons. Thus more electron-hole pairs are produced which can act as carriers of current. Hence, in contrast to a pure metal, the electrical resistance of a semiconductor *decreases* with temperature rise. This is one way of distinguishing between a pure metal and a semiconductor.

P- and N-type Semiconductors

By 'doping' a semiconductor with a tiny amount of impurity such as one part in a million, a considerable increase can be made to the number of charge carriers.

Arsenic atoms, for example, have five electrons in their outermost or valence band. When an atom of arsenic is added to a germanium crystal, the atom settles in a lattice site with four of its electrons shared with neighbouring germanium atoms. Fig. 41.19 (i). The fifth electron

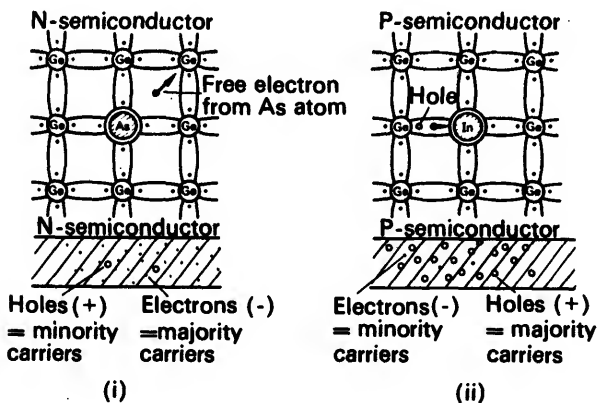


FIG. 41.19. N- and P- semiconductors.

may thus become free to wander through the crystal. Since an impurity atom may provide one free electron, an enormous increase occurs in the number of electron carriers. The impure semiconductor is called an 'n-type semiconductor' or *n-semiconductor*, where 'n' represents the negative charge on an electron. Thus the *majority carriers* in an n-semiconductor are electrons. Positive charges or holes are also present in the n-semiconductor. These are thermally generated, as previously explained, and since they are relatively few they are called

the *minority carriers*. The impurity (arsenic) atoms are called *donors* because they donate electrons as carriers.

P-semiconductors are made by adding foreign atoms which are trivalent to pure germanium or silicon. Examples are boron or indium. In this case the reverse happens to that previously described. Each trivalent atom at a lattice site attracts an electron from a neighbouring atom, thereby completing the four valence bonds and forming a hole in the neighbouring atom. Fig. 41.19 (ii). In this way an enormous increase occurs in the number of holes. Thus in a p-semiconductor, the majority carriers are holes or positive charges. The minority carriers are electrons, negative charges, which are thermally generated. The impurity atoms are called *acceptors* in this case because each 'accepts' an electron when the atom is introduced into the crystal.

Summarizing: In a n-semiconductor, conduction is due mainly to negative charges or electrons, with positive charges (holes) as minority carriers. In a p-semiconductor, conduction is due mainly to positive charges or holes, with negative charges (electrons) as minority carriers.

P-N Junction

By a special manufacturing process, p- and n-semiconductors can be melted so that a boundary or *junction* is formed between them. This junction is extremely thin and of the order 10^{-4} cm. It is called a *p-n junction*. Fig. 41.20 (i). When a scent bottle is opened, the high

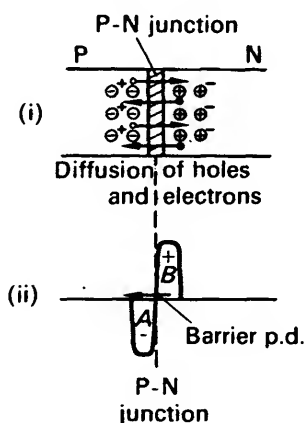


FIG. 41.20. P-n junction and barrier p.d.

concentration of scent molecules in the bottle causes the molecules to diffuse into the air. In the same way, the high concentration of holes (positive charges) on one side of a p-n junction, and the high concentration of electrons on the other side, causes the two carriers to diffuse respectively to the other side of the junction, as shown. The electrons which move to the p-semiconductor side recombine with holes there. These holes therefore disappear, and an excess negative charge A appears on this side. Fig. 41.20 (ii).

In a similar way, an excess positive charge B builds up in the n -semiconductor when holes diffuse across the junction. Together with the negative charge A on the p -side, an e.m.f. or p.d. is produced which opposes the diffusion of charges across the junction. This is called a *barrier p.d.* and when the flow ceases it has a magnitude of a few tenths of a volt.

Junction Diode

When a battery B , with an e.m.f. greater than the barrier p.d., is joined with its positive pole to the p -semiconductor, P , and its negative pole to the n -semiconductor, N , p -charges (holes) are urged across the p - n junction from P to N and n -charges (electrons) from N to P . Fig. 41.21 (i). Thus an appreciable current is obtained. The p - n junction is now said to be *forward-biased*, and when the applied p.d. is increased, the current increases.

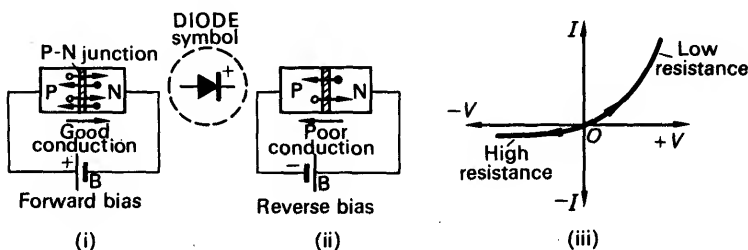


FIG. 41.21. Junction diode characteristic.

When the poles of the battery are reversed, only a very small current flows. Fig. 41.21 (ii). In this case the p - n junction is said to be *reverse-biased*. This time only the minority carriers, negative charges in the p -semiconductor and positive charges in the n -semiconductor, are urged across the p - n junction by the battery. Since the minority carriers are thermally-generated, the magnitude of the reverse current depends only on the temperature of the semiconductors.

It can now be seen that the p - n junction acts as a *rectifier*. It has a low resistance for one direction of p.d. and a high resistance for the opposite direction, as shown by the characteristic curve in Fig. 41.21 (iii). It is therefore called a *junction diode*. The junction diode has advantages over a diode valve; for example, it needs only a low voltage battery B to function; it does not need time to warm up; it is less bulky, and it is cheaper to manufacture in large numbers. On this account, the junction diode has replaced the diode valve in receivers.

Zener Diode

When the reverse bias or p.d. is increased across a p - n junction, a large increase in current is suddenly obtained at a voltage Z . Fig. 41.22 (i). This is called the *Zener effect*, after the discoverer. It is partly due to the high electric field which exists across the narrow p - n junction at the breakdown or Zener voltage Z , which drags more electrons from

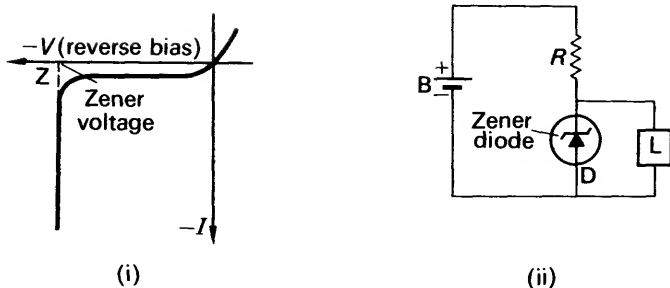


FIG. 41.22. Zener diode and voltage regulation.

their atoms and thus increases considerably the number of electron-hole pairs. Ionization by collision also contributes to the increase in carriers.

Zener diodes are used as voltage regulators or stabilizers in circuits. In Fig. 41.22 (ii), a suitable diode *D* is placed across a circuit *L*. Although the battery supply *B* may fluctuate, and produce changes of current in *L* and *D*, if *R* is suitably chosen, the voltage across *D* remains practically constant over a reverse current range of tens of milliamperes at the Zener voltage. See Fig. 41.22 (i). The voltage across *L* thus remains stable.

The Transistor

The junction diode is a component which can only rectify. The *transistor* is a more useful component; it is a *current amplifier*. A transistor is made from three layers of p- and n-semiconductors. They are called respectively the *emitter* (E), *base* (B) and *collector* (C). Fig. 41.23 (i) illustrates a *p-n-p* transistor, with electrodes connected to the

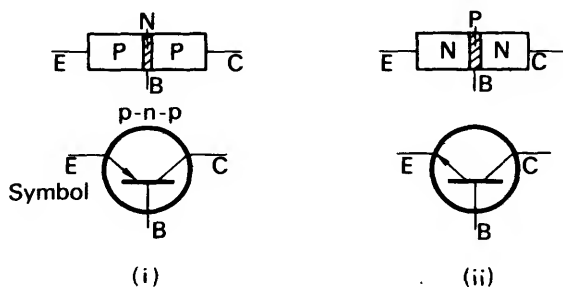


FIG. 41.23. Transistors and symbols.

respective three layers. In a *n-p-n* transistor, the emitter is n-type, the base is p-type and the collector is n-type. Fig. 41.23 (ii). The base is deliberately made very thin in manufacture. The transistor, like the triode valve, is thus a three-terminal device.

Fig. 41.23 shows the circuit symbols for p-n-p and n-p-n transistors. In an actual transistor, the collector terminal is displaced more than the others for recognition or has a dot near it.

Common-Base (C-B) Arrangement

The transistor may be regarded as two p-n junctions back-to-back. Fig. 41.24 (i) shows batteries correctly connected to a p-n-p transistor. The emitter-base is forward-biased; the collector-base is reverse-biased; and the base is common. This is called the *common-base (C-B)* mode of using a transistor. Note carefully the polarities of the two batteries. The positive pole of the supply voltage X is joined to the emitter E but the *negative* pole of the supply voltage Y is joined to the collector C.

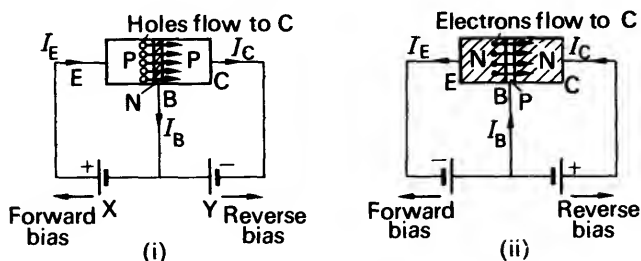


FIG. 41.24. Transistor action.

If batteries are connected the wrong way round to a transistor the latter may be seriously damaged. In the case of a n-p-n transistor, therefore, the negative pole of one battery is joined to the emitter and the positive pole of the other is joined to the collector. Fig. 41.24 (ii).

Consider Fig. 41.24 (i). Here the emitter-base is forward biased by X, so that positive charges or holes flow across the junction from E to the base B. The base is so thin, however, that the great majority of the holes are urged across the base to the collector by the battery Y. Thus a current I_C flows in the collector circuit. The remainder of the holes flow in the base circuit, so that a small current I_B is obtained here. From Kirchhoff's first law, it follows that, if I_E is the emitter current,

$$I_E = I_C + I_B.$$

Typical values for a.f. amplifier transistors are: $I_E = 1.0$ mA, $I_C = 0.98$ mA, $I_B = 0.02$ mA.

Although the action of n-p-n transistors is similar in principle to p-n-p transistors, the carriers of the current in the former case are mainly electrons and in the latter case holes. Electrons are more speedy carriers than holes (p. 1024). Thus n-p-n transistors are used in high-frequency circuits, where the carriers are required to respond very quickly to signals.

Common-Base Characteristics

The behaviour of a particular transistor in the common-base arrangement can be obtained from its characteristic curves. Fig. 41.25 shows a circuit for determining the curves. V_{CC} represents the supply voltage, for example 9 V; A_1 , A_2 are current measuring instruments; V are voltmeters; and the two potentiometers of 1 megohm and 50

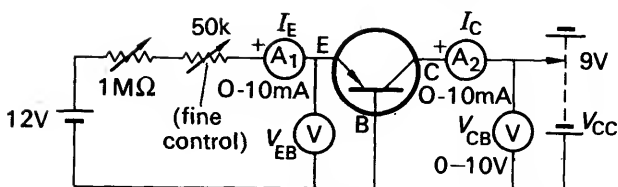


FIG. 41.25. Common-base characteristics investigation.

kilohm are used to vary the input or emitter current, I_E . The more important curves are:

(1) *Output characteristics* (I_C v. V_{CB} , with I_E constant), (2) *input characteristics* (I_E v. V_{EB} , with V_{CB} constant), (3) *transfer characteristics* (I_C v. I_E , with V_{CB} constant).

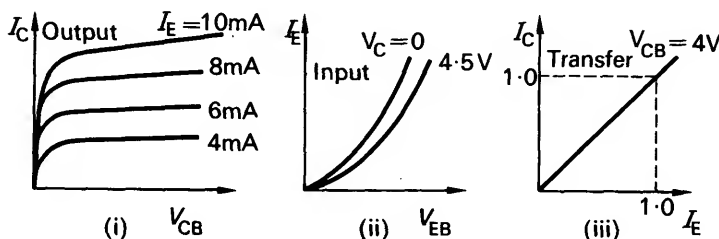


FIG. 41.26. C-B characteristics.

Typical results are shown in Fig. 41.26 (i), (ii), (iii). The flat output characteristics in Fig. 41.26 (i) show that the *output resistance*, $\Delta V_{CB}/\Delta I_C$, is high. The *input resistance*, $\Delta V_{EB}/\Delta I_E$, varies with the slope of the curve in Fig. 41.26 (ii) and is generally low. From the straight line graph of the transfer characteristic in Fig. 41.26 (iii), it follows that a linear relation exists between I_C and I_E .

Common-Emitter (C-E) Arrangement

The slope of the transfer characteristic in Fig. 41.26 (iii) provides the current gain of the transistor, $\Delta I_C/\Delta I_E$. It is practically 1, showing that the common-base arrangement is unsuitable for current amplification. Now in a typical transistor, as already seen, $I_C = 0.98$ mA and $I_B = 0.02$ mA. Thus I_C is 49 times as large as I_B and a similar order of magnitude for current gain occurs with changes in I_B . On this account the common-emitter (or grounded-emitter) arrangement is widely used in a.f. amplifiers.

Fig. 41.27 shows a circuit for determining the output characteristics,

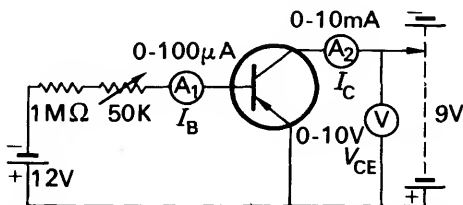


FIG. 41.27. Common-emitter characteristics investigation.

input characteristics and transfer characteristic in the common-emitter (C-E) arrangement. The results are shown in Fig. 41.28 (i), (ii), (iii).

Output characteristic. Since the knee of the curve occurs at a low voltage of the order of 1 V, only low battery supply voltages are needed to operate a transistor in the linear region beyond the knee. This is an advantage of the transistor compared with the valve. Further, the small slope of the straight line shows that the output resistance is high. Thus although the load in the collector circuit may vary, the collector current is constant for a given alternating input or base current. Hence the transistor can be considered as a constant *current generator* in circuitry, whereas the triode valve is treated as a constant *voltage generator* with a given input (p. 1017).

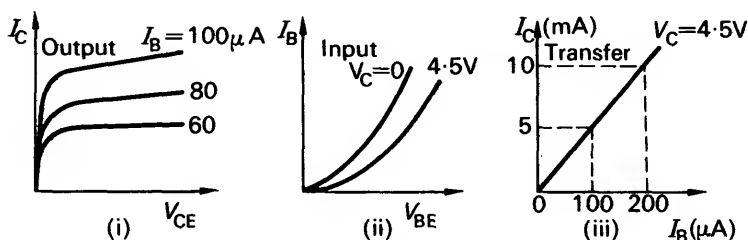


FIG. 41.28. C-E characteristics.

Transfer characteristic. The output current I_C varies fairly linearly with the input current I_B . The current gain, denoted by β , is the ratio $\Delta I_C / \Delta I_B$, V_C constant. From Fig. 41.28 (iii), $\beta = (10 - 5) \text{ mA} / (200 - 100) \mu\text{A} = 50$. The reader should note that 'current amplification' usually refers to variations in current in amplifier circuit analysis. The ratio I_C / I_B provides the d.c. current amplification.

Input characteristic. The input resistance, r_i , is the ratio $\Delta V_{BE} / \Delta I_B$ with V_C constant. It varies at different points of the curve and has a medium value such as 1000Ω or $1 \text{ k}\Omega$.

Relation between Current Gain in C-E and C-B Arrangements

In the C-B arrangement, the current gain is denoted by α and is the ratio $\Delta I_C / \Delta I_E$. In the C-E arrangement, the current gain is denoted by β and is the ratio $\Delta I_C / \Delta I_B$. Now from p. 1029, it is always true that $I_E = I_C + I_B$. Hence $\Delta I_E = \Delta I_C + \Delta I_B$. Using $\Delta I_C / \Delta I_B = \beta$, then $\Delta I_B = \Delta I_C / \beta$. Thus, by substitution for ΔI_B in $\Delta I_E = \Delta I_C + \Delta I_B$,

$$\begin{aligned} \Delta I_E &= \Delta I_C + \frac{\Delta I_C}{\beta} \\ \therefore \frac{\Delta I_E}{\Delta I_C} &= \frac{1}{\alpha} = 1 + \frac{1}{\beta} \\ \text{Simplifying,} \quad \therefore \beta &= \frac{\alpha}{1 - \alpha} \end{aligned} \quad (1)$$

If $\alpha = 0.98$, then $\beta = 0.98 / 0.02 = 49$ from (1).

Leakage Current

When the base current I_B is zero, some current still flows in the collector circuit in the common-emitter arrangement. This is due to the minority carriers present in the collector-base part of the transistor, which is reverse-biased. The collector current when I_B is zero is denoted by I_{CE0} and is called the *leakage current*.

In the common-base arrangement, the leakage current obtained when I_E is zero is denoted by I_{CB0} . This is also due to minority carriers in the collector-base, which is reverse-biased. Thus the leakage current flows when a transistor is in the C-E or C-B arrangement.

Since the current gain β in the C-E arrangement is the ratio $\Delta I_C / \Delta I_B$, with the usual notation, it follows that, generally,

$$I_C = \beta I_B + I_{CE0} \quad . \quad . \quad . \quad (1)$$

Similarly, in the C-B arrangement,

$$I_C = \alpha I_E + I_{CB0} \quad . \quad . \quad . \quad (2)$$

Now any change in I_{CE0} or minority carriers is magnified β times in the C-E arrangement, since I_{CE0} also flows in the base-emitter circuit when the transistor is operating. A temperature change from 25°C to 45°C, which would increase the current I_{CE0} by 10 μA say, would thus be amplified to about $49 \times 10 \mu\text{A}$ or 490 μA , if β is 49. This increase in current, nearly 0.5 mA, would have a considerable effect on the output in the collector circuit, and it would lead to a distorted output, for example.

On the other hand, $\alpha = 0.98$ for the same transistor. Thus in the C-B arrangement, a similar temperature rise and current increase of 10 μA would produce a change in $0.98 \times 10 \mu\text{A}$, or nearly 10 μA , in the output or collector circuit. This is only a very small change compared to the C-E case. On this account the C-E arrangement, which is very sensitive to temperature change, *must* be stabilized for excessive temperature rise. Silicon transistors are much less sensitive to temperature change than germanium transistors and are hence becoming used more widely.

Simple C-E Amplifier Circuit

Fig. 41.29 shows a p-n-p transistor in a simple or basic C-E arrangement. It uses one battery supply, V_{CC} . A load, R_L , is placed in the collector or output circuit. A resistor R provides the necessary bias, V_{BE} .

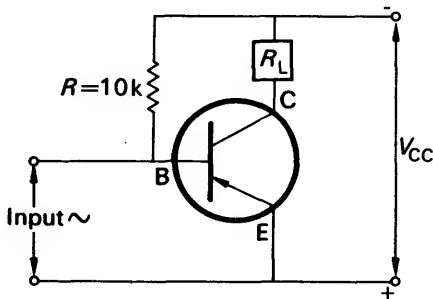


FIG. 41.29. Simple amplifier.

for the base-emitter circuit. The base-emitter is then forward-biased but the collector-base is reverse-biased, that is, the potential of B is negative relative to E but positive relative to C.

Suppose a small signal is applied, so that the base current changes by an amount ΔI_B . Then I_C changes by $\beta \Delta I_B$.

$$\begin{aligned} \therefore \text{voltage gain} &= \frac{\text{output voltage}}{\text{input voltage}} = \frac{\beta \Delta I_B \cdot R_L}{V_i} \\ &= \frac{\beta \Delta I_B \cdot R_L}{\Delta I_B \cdot r_i}, \end{aligned}$$

where r_i is the *input resistance* or resistance to a.c. between base-emitter.

$$\therefore \text{voltage gain} = \frac{\beta \cdot R_L}{r_i}.$$

If $\beta = 49$, $R_L = 4700 \, \Omega$, $r_i = 1000 \, \Omega$, the voltage gain $= 49 \times 4.7 = 230$ (approx.).

C-E Amplifier Circuit

In practice, Fig. 41.29 is unsuitable as an amplifier circuit since there is no arrangement for temperature stabilization. A more reliable C-E a.f. amplifier circuit is shown in Fig. 41.30. Its principal features are:

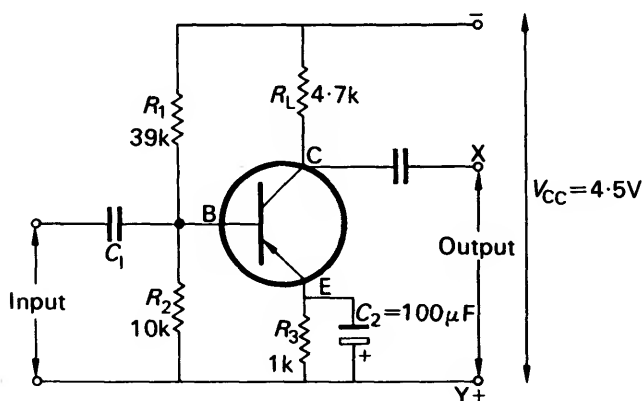


FIG. 41.30. Amplifier circuit.

- (i) a potential divider arrangement, R_1 , R_2 , which provides the necessary base-bias;
- (ii) a load R_L which produces the output across X, Y;
- (iii) a capacitor C_1 which isolates the d.c. component in the input signal from the circuit;
- (iv) a large capacitor C_2 across a resistor R_3 , which prevents undesirable feedback of the amplified signal to the base-emitter circuit;

- (v) an emitter resistance R_3 , which stabilizes the circuit for excessive temperature rise. Thus if the collector current rises, the current through R_3 increases. This lowers the p.d. between E and B, so that the collector current is automatically lowered.

Transistor Oscillator Circuit

Like the triode valve, a transistor can be arranged to provide 'positive feedback' to an oscillatory (L-C) circuit. Oscillations in the L-C circuit can thus be maintained, as explained on p. 1020.

Fig. 41.31 shows one form of transistor oscillator circuit. Its main features are:

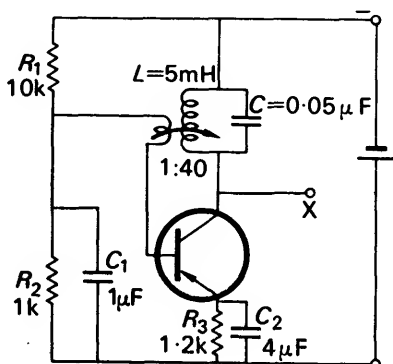


FIG. 41.31. Transistor oscillator.

- (i) a coil-capacitor, L - C , load in the collector circuit;
- (ii) positive feedback through the coil L_1 to maintain the oscillations in the L - C circuit;
- (iii) a potential divider arrangement, R_1 , R_2 , to provide the necessary base bias;
- (iv) an emitter resistor R_3 to stabilize the circuit for excessive temperature rise;
- (v) large capacitors C_1 and C_2 across R_2 and R_3 respectively, which prevent undesirable feedback to the base circuit.

Approximately, the frequency of oscillation is given by $f = 1/2\pi\sqrt{LC}$, in this case an audio-frequency. Other frequencies may be obtained by changing the magnitude of C .

Thermistor

There are numerous semiconductor devices other than the transistor. A *thermistor* is a heat-sensitive resistor usually made from semiconductor materials which have a high negative temperature coefficient of resistance. Its resistance thus decreases appreciably with temperature rise.

One use of a thermistor is to safeguard against current surges in circuits where this could be harmful, for example, in a circuit where the heaters of radio valves are in series. A thermistor, T , is included in the circuit, as shown (Fig.

41.32). When the supply voltage is switched on, the thermistor has a high resistance at first because it is cold. It thus limits the current to a moderate value. As it warms up, the thermistor resistance drops appreciably and an increased current then flows through the heaters. Thermistors are also used in transistor receiver circuits to compensate for excessive rise in collector current.

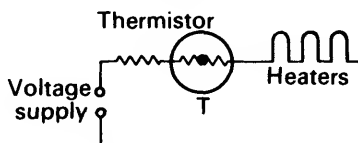


FIG. 41.32. Use of Thermistor.

Phototransistor

A *photodiode* is a junction diode sensitive to light. When the diode is reverse-biased, minority carriers flow in the circuit and constitute a so-called 'dark' current. If the junction of the diode is now illuminated, the light energy produces more electron-hole pairs, which are then swept across the junction. The increased current which flows is the 'light' current.

A *phototransistor* is a transistor sensitive to light in which the base is usually left disconnected. When light falls on the emitter side, more electron-hole pairs are produced in the base. This is amplified by transistor action, and a larger collector current is obtained. In principle the phototransistor is a photodiode plus amplifier.

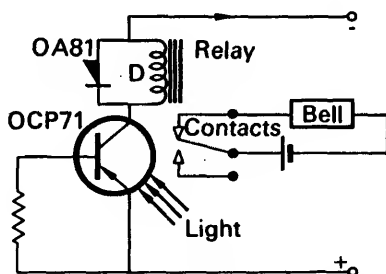


FIG. 41.33. Phototransistor operating relay.

Fig. 41.33 shows a circuit in which a Mullard phototransistor OCP71 is connected in series with a relay coil D and a d.c. supply voltage. When the phototransistor is illuminated, the increase in collector current closes the contacts of a magnetic relay. Current then flows in a circuit connected to the relay, and a bell, for example, may then ring. Fig. 41.33. When the light is switched off, the falling current in the relay coil produces an induced voltage in the same direction as the battery supply. This would raise the collector voltage and prevent the switch-off at the contacts. The diode OA81 across the coil acts as a safeguard. As soon as the rising induced voltage becomes equal to the battery voltage the diode conducts, and prevents any further rise in collector voltage.

EXERCISES 41

1. Sketch graphs using the same axes showing how the current through a thermionic diode varies with the d.c. potential difference applied between the anode and filament for two filament temperatures. Explain three special features of the graphs.

What is meant by (a) half-wave rectification, (b) full-wave rectification? Explain with the aid of labelled circuit diagrams how each of these may be achieved using thermionic diodes. (N.)

2. Figure 41.34 represents a simple rectifier circuit. A sinusoidal 50 cycle alternating voltage of peak value 100 V is applied between *D* and *E*, and a load circuit can be connected between the output terminals *X* and *Y*.

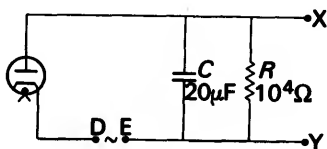


FIG. 41.34.

Draw a diagram showing the variation with time of the p.d. between *X* and *Y* on open circuit. Explain how this is modified when a load is connected across *X* *Y*.

What changes, if any, in the behaviour of the circuit would result from an increase in (a) the temperature of the cathode of the diode, (b) the value of *C*, (c) the value of *R*?

Draw a diagram of a full-wave rectifier-smoother circuit, labelling each component and explaining briefly its purpose. (O.)

3. Give an account of thermionic emission. What analogy exists between thermionic emission and the evaporation of molecules from the surface of a liquid?

Show how the introduction of the grid in a triode enables it to be used as a voltage amplifier. Draw a circuit diagram to show (i) where the voltage to be amplified is applied to the triode, and (ii) where the amplified voltage is tapped off. (O. & C.)

4. Describe the structure of a diode and describe an experiment to justify the term 'valve'.

Explain how a triode (a) differs in structure and operation from a diode, (b) may be used to amplify small alternating potential differences. (L.)

5. Describe the structure of a triode valve and explain the functions of its component parts.

Draw clearly labelled diagrams to show the arrangements necessary (a) to determine the grid characteristics of a triode for various fixed anode potentials, (b) to maintain electrical oscillations by means of a triode. (L.)

6. For a triode, sketch curves to show (a) the form of the anode current/grid voltage static characteristics, (b) the form of the anode current/anode voltage static characteristics. How may the amplification factor of the valve be deduced from these curves?

Explain, with the aid of a circuit diagram and with reference to the static characteristics, how the triode may be used to amplify a small alternating voltage. (N.)

7. Describe the essential features of a triode valve and comment on any *one* feature of its construction which you consider of special importance.

Describe *briefly*, with the aid of a diagram, how you would investigate the variation of anode current with anode potential, the grid potential being constant and negative to the filament. Sketch the curve you would expect to obtain.

A triode valve is to be used to amplify a direct current of 10^{-7} amps flowing in a circuit incorporating a resistance of 10^5 ohms. The valve has a mutual conductance of 2 milliamp volt $^{-1}$ and an anode slope resistance (impedance) of 10^6 ohms. Draw a diagram of a suitable circuit and calculate the current amplification. (N.)

8. What is meant by *thermionic emission*? Describe how this phenomenon is used in the action of a radio valve and give some other use to which it is put.

Show how a triode can be used (a) to detect, (b) to amplify, radio signals. (L.)

9. Give a brief description of the construction of a high-vacuum diode. Draw a graph which shows the variation of the current through such a diode with the potential difference across it, and account for the main features of the curve.

Describe how the introduction of a third electrode, the grid, makes possible (a) control of the current which reaches the anode, and (b) the amplification of a voltage. (O. & C.)

10. Describe with the aid of a circuit diagram how a triode valve can be used as an oscillator. What factors determine (a) the frequency, and (b) the amplitude of the oscillation?

A certain triode valve has an amplification factor of 100 and a mutual conductance of 2.5 mA V $^{-1}$. For reasons of stability, the makers recommend that a resistance of not more than 1.0 megohm shall be connected between grid and cathode. If the valve is used as an amplifier, calculate the maximum possible value of

$$\frac{\text{Power delivered into load resistance}}{\text{Power delivered into the input terminals of the amplifier}} \quad (\text{C.})$$

11. Draw a clear labelled diagram showing the structure of a cathode ray tube.

The potential difference between cathode and anode of a cathode ray tube is 500 volts. The tube is set up with its axis along the direction of the earth's resultant magnetic field and the spot is focussed on the screen which is 15.0 cm from the anode. On rotating the tube about a horizontal axis to a position at right angles to the earth's resultant field the spot is deflected through 0.75 cm. Find a value for e/m for an electron. (Assume the magnetic induction in the earth's magnetic field to be 0.50×10^{-4} Wb m $^{-2}$.) (L.)

12. Describe an experiment to determine the deflection sensitivity of a cathode-ray tube in volts per cm.

Give an account of any experiment you have performed, or seen performed, in which a cathode ray oscilloscope is used to obtain information. Explain the purpose of the experiment, and the nature of the information obtained from the oscillograph. (C.)

13. Draw a sketch to show the essential parts of a cathode ray oscillograph having electrostatic deflection.

With the help of your sketch explain how in a cathode ray oscillograph: (a) the electrons are produced; (b) the electrons are focused; (c) the spot is made visible; (d) the brightness of the spot is controlled.

What is meant by stating that a cathode ray oscillograph is fitted with a linear time base of variable frequency. (N.)

14. Explain what is meant by (a) a linear time-base, (b) a sinusoidal time-base, in a cathode ray oscilloscope.

The X and Y deflection sensitivities of a cathode ray oscilloscope are each 5 volt cm^{-1} . A sinusoidal potential difference alternating at 50 Hz and of r.m.s. value 20 volts is applied to the Y plates of the instrument. A potential difference of the same form and frequency but of r.m.s. value 10 volts is simultaneously applied to the X plates. Sketch and explain the pattern seen on the oscilloscope when the potential differences are (a) in phase, (b) 90° out of phase. Indicate the appropriate dimensions on your sketches. (N)

15. Give an account of the cathode ray tube, and explain how the 'brilliance', 'focus' and 'shift' controls operate.

How could you use the cathode ray tube to measure d.c. voltages?

It is proposed to use a cathode ray tube to study the amplification of a thermionic triode by applying a sinusoidal alternating voltage V_C between the grid and the cathode and comparing this with the alternating voltage V developed across a resistor in the anode circuit. Draw a suitable circuit diagram, showing how the voltages V_C and V may be applied in turn across the Y plates of the cathode ray tube.

How would you expect the ratio V/V_C to depend on the value of the anode resistor? (O)

16. Explain, with special reference to a cathode ray tube how a stream of electrons may be produced and its direction and intensity controlled.

The focussing of a C.R.O. may be done by using two cylindrical anodes at different potentials. Sketch the lines of force between two such anodes, and use your sketch to explain how a diverging beam of electrons may be made converging.

100 volts a.c. applied to the Y -plates of a C.R.O. give a sinusoidal trace which measures 6 cm vertically (peak to peak). When 10 volts sinusoidal a.c. are applied to the input of a simple triode amplifier and the output from the amplifier is connected to the Y -plates of the C.R.O., the trace produced is as shown in Fig. 41B. Describe the output from the amplifier. How may it be explained in terms of triode characteristics? (N)

Junction Diode. Transistor

17. Explain, with reference to the carriers, the effect of temperature rise on the resistance of a pure metal and on the resistance of a pure semiconductor.

18. Explain what is meant by (i) a p - and a n -semiconductor. (ii) a p - n junction.

19. Draw a sketch of the characteristic of a p - n junction diode. Explain, in terms of the movement of carriers, why the resistance of the diode is low in one direction and high in the reverse direction.

20. Draw a sketch of a p - n - p transistor used in (i) a common-base (CB) and (ii) a common-emitter (CE) arrangement, showing clearly the polarities of the batteries. Explain why the common-emitter arrangement is preferred in an a.f. amplifier circuit.

21. Draw a circuit showing how the collector current-collector voltage and the emitter current-emitter voltage characteristics of a transistor can be found for the common-emitter (CE) arrangement. Sketch the characteristics obtained.

Draw a sketch of a simple CE a.f. amplifier.

22. A transistor in the common-emitter arrangement provides the following results for I_C , collector current, and V_C , collector voltage, for various constant base currents I_B :

V_C (V)	I_C (mA)			
	$I_B = 20 \mu A$	$= 40 \mu A$	$= 60 \mu A$	$= 80 \mu A$
3	0.91	1.60	2.30	3.00
5	0.93	1.70	2.50	3.25
7	0.97	1.85	2.70	3.55
9	1.00	2.05	3.00	4.05

Plot the characteristics, and from them find (i) the current gain at 8 V, (ii) the output resistance for a base current of $40 \mu A$.

23. Explain what is meant by p-type and n-type semiconductors. Describe a p-n junction diode. Draw a graph which shows the variation of the current through such a diode with the potential difference across it, and explain why the diode behaves differently when the potential difference across it is reversed.

Describe the junction transistor. Sketch curves to show the variation of the collector current with the collector-base voltage for various values of the emitter current and explain their form. (O. & C.)

chapter forty-two

Radioactivity. The Nucleus

IN 1896 Becquerel found that a uranium compound affected a photographic plate wrapped in light-proof paper. He called the phenomenon *radioactivity*, and we shall see later that natural radioactivity is due to one or more of three types of radiation emitted from heavy elements such as uranium. These were originally called α -, β - and γ -rays but α - and β -‘rays’ were soon shown to be actually particles.

α - and β -particles and γ -rays all produce ionization as they move through a gas. On average, α -particles produce about 1000 times as many ions per unit length of their path as β -particles, which in turn produce about 1000 times as many ions as γ -rays. There are numerous detectors of ionizing radiations such as α - and β -particles and γ -rays. We begin by describing two detectors used in laboratories.

Geiger-Müller Tube

A Geiger-Muller (GM) tube is widely used for detecting ionizing particles. In one form it consists of an insulated wire A mounted in a thin-walled glass tube B coated with aquadag (a colloidal suspension of graphite) and earthed (Fig. 42.1). A p.d. V of the order of 400 volts is maintained across A, B. When a single ionizing particle enters the

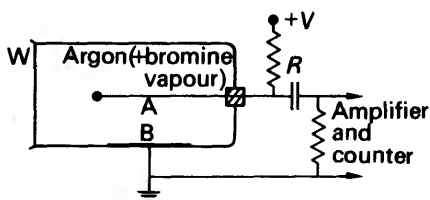


FIG. 42.1. Principle of Geiger-Muller tube.

chamber, a few electrons and ions are produced in the gas. If V is above the breakdown potential of the gas, the number of electrons and ions are multiplied enormously (see p. 998). The electrons are attracted by and move towards A, and the positive ions move towards B. Thus a ‘discharge’ is suddenly obtained between A and B. The current flowing in the high resistance R produces a p.d. which is amplified and passed to a counter, discussed on p. 1041. This registers the passage of an ionizing particle or radiation through the tube.

The discharge persists for a short time, as secondary electrons are emitted from the cathode by the positive ions which arrive there. This would upset the recording of other ionizing particles following fast on

the first one recorded. The air in the tube is therefore replaced by argon mixed with bromine vapour, which has the property of quenching the discharge quickly. Electrical methods are also used for quenching.

Solid State Detector

A *solid state detector*, Fig. 42.2, is made from semiconductors. Basically, it has a p-n junction which is given a small bias in the non-conduction direction (p. 1026). When an energetic ionizing particle such

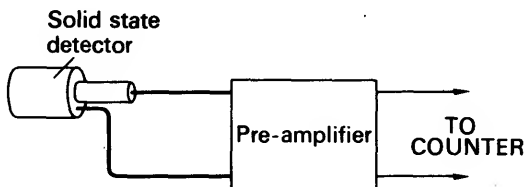


FIG. 42.2. Solid State Detector.

as an α -particle falls on the detector, more electron-hole pairs are created near the junction. These charge carriers move under the influence of the biasing potential and so a pulse of current is produced. The pulse is fed to an amplifier and the output passed to a counter.

The solid state detector is particularly useful for α -particle detection. If the amplifier is specially designed, β -particles and γ -rays of high energy may also be detected. This type of detector can thus be used for all three types of radiation.

Dekatron Counter. Ratemeter

As we have seen, each ionizing particle or radiation produces a pulse voltage in the external circuit of a Geiger-Muller or solid state detector. In order to measure the number of pulses from the detectors, some form of counter must be used.

A *dekatron counter* consists of two or more dekatron tubes, each containing a glow or discharge which can move round a circular scale graduated in numbers 0-9, together with a mechanical counter (Fig. 42.3). Each impulse causes the discharge in the first tube, which

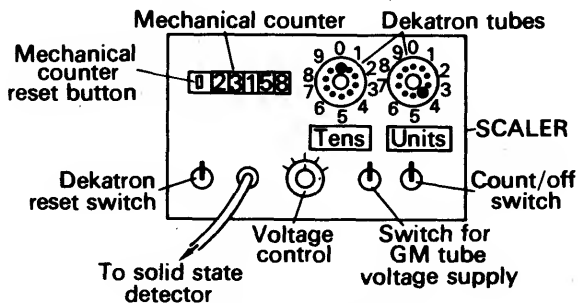


FIG. 42.3. Dekatron counter.

counts units, to advance one digit. The circuit is designed so that on the tenth pulse, which returns the first counter to zero, a pulse is sent to the second tube. The glow here then moves on one place. The second tube thus counts the number of tens of pulses. After ten pulses are sent to the second tube, corresponding to a count of 100, the output pulse from the second tube is fed to the mechanical counter. This, therefore, registers the hundreds, thousands and so on. Dekatron tubes are used in radioactive experiments because they can respond to a rate of about 1000 counts per second. This is greatly in excess of the count rate possible with a mechanical counter.

In contrast to a scaler, which counts the actual number of pulses, a *ratemeter* is a device which provides directly the average number of pulses per second or *count rate*. The principle is shown in Fig. 42.4. The

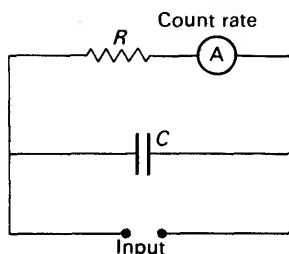


FIG. 42.4. Principle of a Ratemeter.

pulses received are passed to a capacitor C , which then stores the charge. C discharges slowly through a high resistor R and the average discharge current is recorded on a microammeter A . The greater the rate at which the pulses arrive, the greater will be the meter reading. The meter thus records a current which is proportional to the count rate.

If a large capacitor C is used, it will take a long time to charge and the pulses will be averaged over a long time. A switch marked 'time constant' on most ratemeters allows the magnitude of C to be chosen. If a large value of C is used, the capacitor will take a relatively long time to charge and correspondingly it will be a long time before a reading can be taken. The reading obtained, however, will be more accurate since the count rate is then averaged over a longer time (see below). For high accuracy, a small value of C may be used only if the count rate is very high.

Errors in Counting Experiments

Radioactive decay is random in nature (p. 1048). If the count rate is high, it is not necessary to wait so long before readings are obtained which vary relatively slightly from each other. If the count rate is low, successive counts will have larger percentage differences from each other, unless a much longer counting time is employed.

The accuracy of a count does not depend on the time involved but on the *total count obtained*. If N counts are received, the statistics of random processes show that this is subject to a statistical error of

$\pm \sqrt{N}$. The proof of this is beyond the scope of this book. The percentage error is thus

$$\frac{\sqrt{N}}{N} \times 100 = \frac{100}{\sqrt{N}}\%$$

If 10 per cent accuracy is required, $\sqrt{N} = 10$ and hence $N = 100$. Thus 100 counts must be obtained. If the counts are arriving at about 10 every second, it will be necessary to wait for 10 seconds to obtain a count of 100 and so achieve 10 per cent accuracy. Thus a ratemeter circuit must be arranged with a time constant (CR) of 10 seconds, so that an average is obtained over this time. If, however, the counts are arriving at a rate of 1000 per second on average, it will be necessary to wait only 1/10th second to achieve 10 per cent accuracy. Thus the 1 second time constant scale on the ratemeter will be more than adequate.

Existence of α -, β -particles and γ -rays

The existence of different ionizing particles or radiations from radioactive substances can be shown by an absorption experiment, using a counter or ratemeter.

A radium source S , producing α - and β -particles and γ -rays, is placed at a fixed small distance from a solid state detector A and sensitive low-noise pre-amplifier, which is connected to a counter C (or ratemeter) (Fig. 42.5 (i)). Foils of increasing thickness are placed over the

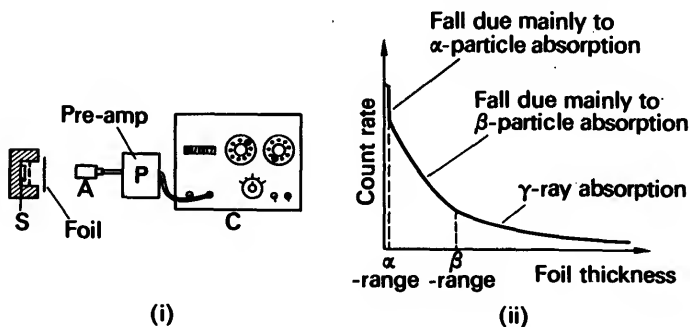


FIG. 42.5. Existence of α -, β -particles and γ -rays.

source, starting with *very* thin foils of paper and then aluminium foils. For each thickness the count rate is measured. A graph of count rate against thickness of foil is then plotted.

The resulting curve, Fig. 42.5 (ii), shows three distinct portions. To begin with, α , β and γ all pass through the very thin foils such as paper. After a particular thickness the α -particles are absorbed, and beyond this point the curve does not then fall off with distance as quickly. A similar change takes place when the β -particles are all absorbed at a particular thickness of aluminium plate, leaving another radiation, the γ -rays. This straightforward experiment shows the existence of three different types of radiation, α - and β -particles and γ -rays.

Alpha-particles

It is found that α -particles have a fairly definite range in air at atmospheric pressure. This can be shown by slowly increasing the distance between a pure α -source and a detector. The count rate is observed to fall rapidly to zero at a separation greater than a particular value, which is called the 'range' of the α -particles. The range depends on the source and on the air pressure.

Using the apparatus of Fig. 42.6, it can be shown the α -particles are *positively* charged. When there is no magnetic field, the solid state detector is placed so that

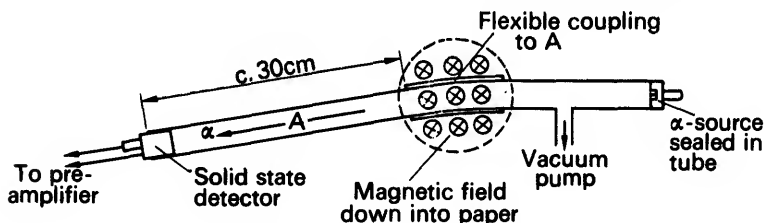


FIG. 42.6. Charge on an α -particle.

the tube A is horizontal in order to get the greatest count. When the magnetic field is applied, the detector has to be moved *downwards* in order to get the greatest count. This shows that the α -particles are deflected by a small amount downwards. By applying Fleming's left-hand rule, we find that particles are *positively* charged. The vacuum pump is needed in the experiment, as the range of α -particles in air at normal pressures is too small.

Nature of α -particle

Lord Rutherford and his collaborators found by deflection experiments that an α -particle had a mass about four times that of a hydrogen atom, and carried a charge $+2e$, where e was the numerical value of the charge on an electron. The atomic weight of helium is about four. It was thus fairly certain that an α -particle was a *helium nucleus*, that is, a helium atom which has lost two electrons.

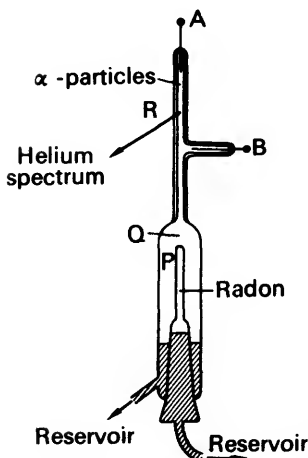


FIG. 42.7. Rutherford and Royds' experiment on α -particles.

In 1909 Rutherford and Royds showed conclusively that α -particles were helium nuclei. Radon, a gas given off by radium which emits α -particles, was collected above mercury in a thin-walled tube P (Fig. 42.7). After several days some of the α -particles passed through P into a surrounding vacuum Q, and in about a week, the space in Q was reduced in volume by raising mercury reservoirs. A gas was collected in a capillary tube R at the top of Q. A high voltage from an induction coil was then connected to electrodes at A and B, and the spectrum of the discharge was observed to be exactly the same as the characteristic spectrum of helium.

Beta-particles and Gamma-rays

By deflecting β -particles with perpendicular magnetic and electric fields, their charge-mass ratio could be estimated. This is similar to Thomson's experiment, p. 1003. These experiments showed that *β -particles are electrons moving at high speeds*. Generally, β -particles have a greater penetrating power of materials than α -particles. They also have a greater range in air than α -particles, since their ionization of air is relatively smaller, but their path is not so well defined.

Using a Ticonal bar magnet, it can be shown that β -particles are strongly deflected by a magnetic field. The direction of the deflection corresponds to a stream of *negatively-charged* particles, that is, opposite to the deflection of α -particles in the same field. This is consistent with the idea that β -particles are usually fast-moving electrons.

The nature of γ -rays was shown by experiments with crystals. Diffraction phenomena are obtained in this case, which suggest that *γ -rays are electromagnetic waves* (compare X-rays, p. 1067). Measurement of their wavelengths, by special techniques with crystals, show they are shorter than the wavelengths of X-rays and of the order 10^{-9} cm. γ -rays can penetrate large thicknesses of metals, but they have far less ionizing power in gases than β -particles.

If a beam of γ -rays are allowed to pass through a very strong magnetic field no deflection is observed. This is consistent with the fact that γ -rays are electromagnetic waves and carry no charge.

Inverse-square Law for γ -rays

If γ -rays are a form of electromagnetic radiation and undergo negligible absorption in air, then the intensity I should vary inversely as the square of the distance between the source and the detector. The apparatus shown in Fig. 42.8 can be used to investigate if this is the case. A pure γ -source is placed at a suitable distance from a GM tube connected to a scaler, and I will then be proportional to the count rate C .

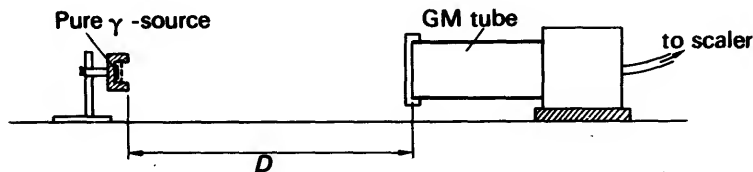


FIG. 42.8. Inverse square law for γ -rays.

Suppose D is the measured distance from a fixed point on the γ -source support to the front of the GM tube. To obtain the true distance from the source to the region of gas inside the tube where ionization occurs, we need to add an unknown but constant distance h to D . Then, assuming an inverse-square law, $I \propto 1/(D+h)^2$. Thus

$$D+h \propto \frac{1}{\sqrt{I}} \propto \frac{1}{\sqrt{C}}$$

A graph of $1/\sqrt{C}$ is therefore plotted against D for varying values of D . If the inverse-square law is true, a straight line graph is obtained which has an intercept on the D -axis of $-h$. Note that if I is plotted against $1/D^2$ and h is not zero, a straight line graph is *not* obtained from the relation $I \propto 1/(D+h)^2$. Consequently we need to plot D against $1/\sqrt{I}$.

If a pure β -source is substituted for the γ -source and the experiment is repeated, a straight-line graph is not obtained. The absorption of β -particles in air is thus appreciable compared with γ -rays.

Half-life Period

Radioactivity, or the emission of α - or β -particles and γ -rays, is due to disintegrating nuclei of atoms (p. 1052). The disintegrations obey the statistical law of chance. Thus although we can not tell which particular atom is likely to disintegrate next, the number of atoms disintegrating per second, dN/dt , is directly proportional to the number of atoms, N , present at that instant. Hence:

$$\frac{dN}{dt} = -\lambda N;$$

where λ is a constant characteristic of the atom concerned called the *radioactivity decay constant*. Thus, if N_0 is the number of radioactive atoms present at a time $t = 0$, and N is the number at the end of a time t , we have, by integration,

$$\begin{aligned} \int_{N_0}^N \frac{dN}{N} &= -\lambda \int_0^t dt. \\ \therefore \left[\log_e N \right]_{N_0}^N &= -\lambda t. \\ \therefore N &= N_0 e^{-\lambda t} \end{aligned} \quad (i)$$

Thus the number N of radioactive atoms left decreases exponentially with the time t , and this is illustrated in Fig. 42.9.

The *half-life period* $T_{1/2}$ of a radioactive element is defined as the time taken for half the atoms to disintegrate (see Fig. 42.9), that is, in a time $T_{1/2}$ the radioactivity of the element diminishes to half its value. Hence, from (i),

$$\begin{aligned} \frac{N_0}{2} &= N_0 e^{-\lambda t} \\ \therefore T_{1/2} &= \frac{1}{\lambda} \log_e 2 = \frac{0.693}{\lambda}. \end{aligned} \quad (ii)$$

The half-life period varies considerably in a particular radioactive series. In the uranium series shown in the Table on p. 1053, for example, uranium I has a half-life period of the order of 4500 million years, radium has one of about 1600 years, radium F about 138 days, radium B about 27 minutes, and radium C' about 10^{-4} second.

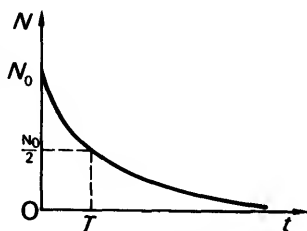


FIG. 42.9. Radioactive decay with time.

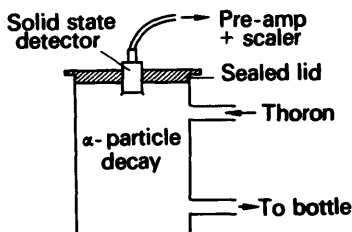


FIG. 42.10. Half life of thoron.

The half-life of thoron, a radioactive gas with a short half-life, can be measured by the apparatus shown in Fig. 42.10. A solid state detector is mounted inside a closed chamber and some thoron gas is passed in from a bottle containing thorium hydroxide which produces the gas. The total count is measured every 30 seconds for about 5 minutes, when counting has virtually stopped, and a graph of total count v. time is then plotted.

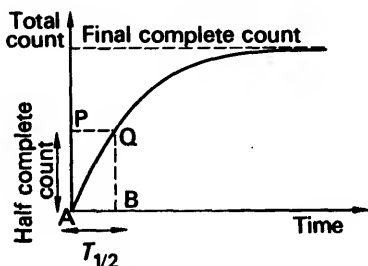


FIG. 42.11. Count in half life experiment.

Fig 42.11 shows the shape of the graph obtained. To find the half-life, a line PQ is drawn at half the final count. Half the atoms have disintegrated in the time AB and hence AB is the half-life, $T_{1/2}$, of thoron. This is read from the axis.

EXAMPLE

At a certain instant, a piece of radioactive material contains 10^{12} atoms. The half-life of the material is 30 days.

(1) Calculate the number of disintegrations in the first second. (2) How long will elapse before 10^4 atoms remain? (3) What is the count rate at this time?

(1) We have

$$N = N_0 e^{-\lambda t}$$

$$\therefore \frac{dN}{dt} = -N_0 \lambda e^{-\lambda t} = -\lambda N.$$

Hence, when $N = 10^{12}$, $\frac{dN}{dt} = -\lambda 10^{12}$.

Now

$$\lambda = \frac{0.693}{T} = \frac{0.693}{30 \times 24 \times 60 \times 60} \text{ s}^{-1}$$

\therefore number of disintegrations per second

$$= \frac{10^{12} \times 0.693}{30 \times 24 \times 60 \times 60} = 2.7 \times 10^5$$

(2) When $N = 10^4$, we have

$$10^4 = 10^{12} e^{-\lambda t}$$

$$\therefore 10^{-8} = e^{-\lambda t}$$

Taking logs to base 10,

$$\therefore \log 10^{-8} = \log_{10} e^{-\lambda t}$$

$$\therefore -8 = -\lambda t \log_{10} e$$

$$\therefore t = \frac{8}{\lambda \log_{10} e} = \frac{8T}{0.693 \log_{10} e}$$

$$= 64 \text{ days (approx.)}$$

(3) Since

$$\frac{dN}{dt} = -\lambda N$$

$$\therefore \text{number of disintegrations per hour} = \frac{0.693}{30 \times 24} \times 10^4$$

$$= 9.6.$$

Wilson's Cloud Chamber

C. T. R. Wilson's *cloud chamber*, invented in 1911, was one of the most useful early inventions for studying radioactivity. It enabled photographs to be obtained of ionizing particles or radiation.

Basically, Wilson's cloud chamber consists of a chamber Y into which saturated water-vapour is introduced (Fig. 42.12). When the pressure is suddenly reduced below a hollow glass piston X, the latter drops down and the air in Y undergoes an adiabatic expansion and cools. The dust nuclei are all carried away after a few expansions by drops forming on them, and then the dust-free air in Y is subjected to a controlled adiabatic expansion of about 1.31 to 1.38 times its original volume. The air is now supersaturated, that is, the vapour pressure is greater than the saturation vapour pressure at the reduced temperature reached but no water-vapour condenses. Simultaneously, the air is exposed to ionizing agents such as α -, β -particles or γ -rays, and water droplets immediately collect round the ions produced which act as centres of formation. The drops are photographed by light scattered from them, and in this way the tracks of ionizing particles or radiation are made visible. Wilson's cloud chamber has proved of immense value in the study of radioactivity and nuclear structure.

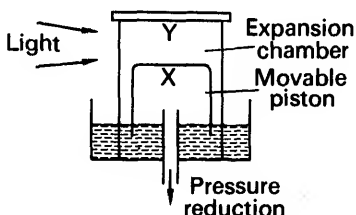


FIG. 42.12. Principle of cloud chamber—Wilson.

The random nature of radioactive decay can be seen by using a Wilson cloud chamber. Particles emitted by a radioactive substance do not appear at equal intervals of time but are sporadic or entirely random. The length of the track of an emitted particle is a measure of its initial energy. The tracks of α -particles are nearly all the same,

showing that the α -particles were all emitted with the same energy. Sometimes two different lengths of tracks are obtained, showing that the α -particles may have one of two energies on emission.

Glaser's Bubble Chamber

In the same way as air can be super-saturated with water vapour, a liquid under pressure can be heated to a temperature higher than that at which boiling normally takes place and is then said to be *superheated*. If the pressure is suddenly released, bubbles may not form in the liquid for perhaps 30 seconds or more. During this quiet period, if ionizing particles or radiation are introduced into the liquid, nuclei are obtained for bubble formation. The liquid quickly evaporates into the bubble, which grows rapidly, and the bubble track when photographed shows the path of the ionizing particle.

Glaser invented the bubble chamber in 1951. It is now widely used in nuclear investigations all over the world, and it is superior to the cloud chamber. The density of the liquid ensures shorter tracks than in air, so that a nuclear collision of interest by a particle will be more likely to take place in a given length of liquid than in the same length of air. Photographs of the tracks are much clearer than those taken in the cloud chamber, and they can be taken more rapidly. In 1963 a 1.5 metre liquid hydrogen bubble chamber was constructed for use at the Rutherford High Energy Laboratory, Didcot, England. High energy protons, accelerated by millions of volts are used to bombard hydrogen nuclei in the chamber. The products of the reaction are bent into a curved track by a very powerful magnetic field, and the appearance and radius of the track then provides information about the nature, momentum or energy of the particles emitted.

Scintillations and Photomultiplier

In the early experiments on radioactivity, Rutherford observed the scintillations produced when an α -particle was incident on a material such as zinc sulphide. This is now utilized in the scintillation photomultiplier, whose principle is illustrated in Fig. 42.13. When an ionizing

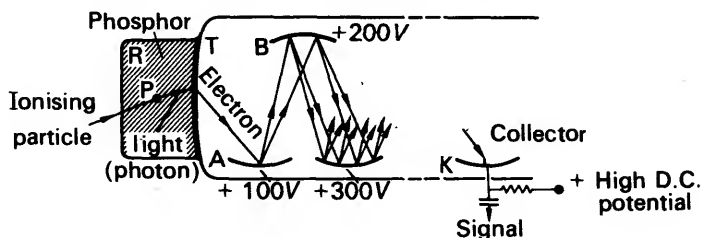


FIG. 42.13. Principle of photomultiplier.

particle strikes the scintillation material or *phosphor* S, the light falls on a photo-sensitive material A and ejects electrons. In one type of tube, these are now focused towards and accelerated to an electrode B, coated with a material which emits secondary electrons four or five

times as numerous as those incident on it. The secondary electrons then strike an electrode C after further acceleration, thus multiplying the number of electrons further, and so on along the tube. A single ionizing particle can produce a million electrons in a photomultiplier tube, and the pulse of current is amplified further and recorded. By choosing a suitable phosphor, scintillation counters can detect electrons and gamma rays, as well as fast neutrons.

Emulsions

Special photographic emulsions have been designed for investigating nuclear reactions. The emulsions are much thicker than those used in ordinary photography, and the concentration of silver bromide in gelatine is many times greater than in ordinary photography. α -particles, protons and neutrons can be detected in specially-prepared emulsions by the track of silver granules produced, which has usually a very short range of the order of a millimetre or less. Consequently, after the plate is developed the track is observed under a high power microscope, or a photomicrograph is made. Nuclear emulsions were particularly useful in investigations of cosmic rays at various altitudes.

THE NUCLEUS

Discovery of Nucleus

In 1909 Geiger and Marsden, at Lord Rutherford's suggestion, investigated the scattering of α -particles by thin films of metal of high atomic weight, such as gold foil. They used a radon tube S in a metal-block as a source of α -particles, and limited the particles to a narrow pencil (Fig. 42.14). The thin metal foil A was placed in the centre of an evacuated vessel, and the scattering of the particles after passing through A was observed on a fluorescent screen B, placed at the focal plane of a microscope M. Scintillations are seen on B whenever it is struck by α -particles.

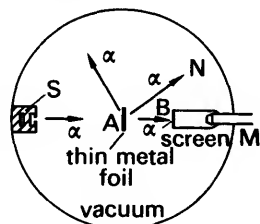


FIG. 42.14. Discovery of nucleus—Geiger and Marsden.

Geiger and Marsden found that α -particles struck B not only in the direction SA, but also when the microscope M was moved round to N and even to P. Thus though the majority of α -particles were scattered through small angles, some particles were scattered through very large angles. Rutherford found this very exciting news. It meant that some α -particles had come into the repulsive field of a highly concentrated positive charge at the heart or centre of the atom, and on the basis of an inverse-square law repulsion he calculated the number of α -particles scattered in a definite direction. The relationship was verified by Geiger and Marsden in subsequent experiments. An atom thus has a *nucleus*, in which all the positive charge and most of its mass is concentrated.

Atomic Mass and Atomic Number

In 1911 Rutherford proposed the basic structure of the atom which is accepted today, and which subsequent experiments by Moseley and others have confirmed. A neutral atom consists of a very tiny nucleus of diameter about 10^{-13} cm which contains practically the whole mass of the atom. The atom is largely empty. If a drop of water was magnified until it reached the size of the earth, the atoms inside would then be only a few metres in diameter and the atomic nucleus would have a diameter of only about 10^{-2} millimetre.

The nucleus of hydrogen is called a *proton*, and it carries a charge of $+e$, where e is the numerical value of the charge on an electron. The helium nucleus has a charge of $+2e$. The nucleus of copper has a charge of $+29e$, and the uranium nucleus carries a charge of $+92e$. Generally, the positive charge on a nucleus is $+Ze$, where Z is the *atomic number* of the element and is defined as the number of protons in the nucleus (see also p. 1072). Under the attractive influence of the positively-charged nucleus, a number of electrons equal to the atomic number move round the nucleus and surround it like a negatively-charged cloud.

Discovery of Protons in Nucleus

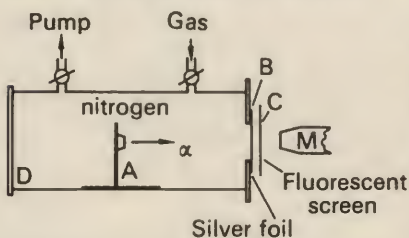


FIG. 42.15. Discovery of protons in the nucleus—Rutherford.



In 1919 Rutherford found that energetic α -particles could penetrate nitrogen atoms and that protons were ejected after the collision. The apparatus used is shown in Fig. 42.15. A source of α -particles, A, was placed in a container D from which all the air had been pumped out and replaced by nitrogen. Silver foil, B, sufficiently thick to stop α -particles, was then placed between A and a fluorescent screen C, and scintillations were observed by a microscope M. The particles which have passed through B were shown to have a similar range, and the same charge, as protons.

Protons were also obtained with the gas fluorine, and with other elements such as the metals sodium and aluminium. It thus became clear that the nuclei of all elements contain

FIG. 42.16. Transmutation of Nitrogen by collision with α -particle. An oxygen nucleus, right-curved track, and a proton, left straight track, are produced.

protons. The number of protons must equal the number of electrons surrounding the nucleus, so that each is equal to the atomic number, Z , of the element. A proton is represented by the symbol, ${}^1_1\text{H}$; the top number denotes the *mass number*, the whole number nearest to the relative atomic mass, and the bottom number the nuclear charge in units of $+e$. The helium nucleus such as an α -particle is represented by ${}^4_2\text{He}$; its mass number is 4 and its nuclear charge is $+2e$, so that the nucleus contains two protons. One of the heaviest nuclei, uranium, can be represented by ${}^{238}_{92}\text{U}$; it has a mass number of 238 and a nuclear charge of $+92e$, so that its nucleus contains 92 protons.

Discovery of Neutron in Nucleus

In 1930 Bothe and Becker found that a very penetrating radiation was produced when α -particles were incident on beryllium. Since the radiation had no charge it was thought to be γ -radiation of very great energy. In 1932 Curie-Joliot placed a block of paraffin-wax in front of the penetrating radiation, and showed that protons of considerable range were ejected from the paraffin-wax. The energy of the radiation could be calculated from the range of the ejected proton, and it was then found to be improbably high.

In 1932 Chadwick measured the velocity of protons and of nitrogen nuclei when they were ejected from materials containing hydrogen and nitrogen by the penetrating radiation. He used polonium, A, as a source of α -particles and the unknown radiation X, obtained by impact with beryllium, B, was then incident on a slab C of paraffin-wax (Fig.

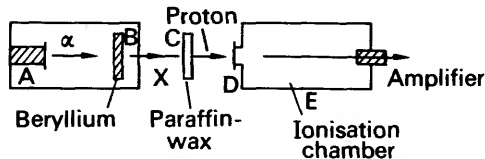


FIG. 42.17. Discovery of the neutron—Chadwick.

42.17). The velocity of the protons emitted from C could be found from their range in air, which was determined by placing various thicknesses of mica, D, in front of an ionization chamber, E, until no effect was produced here. By previous calibration of the thickness of mica in terms of air thickness, the range in air was found.

Chadwick repeated the experiment with a slab of material containing nitrogen in place of paraffin-wax. He then applied the laws of conservation of linear momentum and energy to the respective collisions with the hydrogen and nitrogen atoms, assuming that the unknown radiation was a *particle* carrying no charge and the collisions were elastic. From the equations obtained, he calculated the mass of the particle, and found it to be about the same mass as the proton. Chadwick called the new particle a *neutron*, and it is now considered that all nuclei contain protons and neutrons. The neutron is represented by the symbol ${}_0^1\text{n}$ as it has an atomic mass of 1 and zero charge.

We can now see that a helium nucleus, ${}^4_2\text{He}$, has 2 protons and 2 neutrons, a total mass number of 4 and a total charge of $+2e$. The

sodium nucleus, $^{23}_{11}\text{Na}$, has 11 protons and 12 neutrons. The uranium nucleus, $^{238}_{92}\text{U}$, has 92 protons and 146 neutrons. Generally, a nucleus represented by ^A_ZX has Z protons and $(A - Z)$ neutrons.

Radioactive Disintegration

Naturally occurring radioactive elements such as uranium, actinium and thorium disintegrate to form new elements, and these in turn are unstable and form other elements. Between 1902 and 1909 Rutherford and Soddy made a study of the elements formed from a particular 'parent' element, and the *uranium series* is listed in the table below.

Element	Symbol	Atomic Number	Mass Number	Half-life Period (T)	Particle emitted
Uranium I (U)	UI	92	238	4,500 million years	α
Uranium X ₁ (Th)	UX ₁	90	234	24 days	β, γ
Uranium X ₂ (Pa)	UX ₂	91	234	1.2 minutes	β, γ
Uranium II (U)	UII	92	234	250,000 years	α
Ionium (Th)	Io	90	230	80,000 years	α, γ
Radium	Ra	88	226	1,600 years	α, γ
Radon	Rn	86	222	3.8 days	α
Radium A (Po)	RaA	84	218	3 minutes	α
Radium B (Pb)	RaB	82	214	27 minutes	β, γ
Radium C (Bi)	RaC	83	214	20 minutes	β, γ
Radium C' (Po)	RaC'	84	214	1.6×10^{-4} seconds	α
Radium C'' (Tl)	RaC''	81	210	1.3 minutes	β
Radium D (Pb)	RaD	82	210	19 years	β, γ
Radium E (Bi)	RaE	83	210	5 days	β
Radium F (Po)	RaF	84	210	138 days	α, γ
Lead	Pb	82	206	(stable)	

The new element formed after disintegration can be identified by considering the particles emitted from the nucleus of the parent atom. An α -particle, a helium nucleus, has a charge of $+2e$ and a mass number 4. Uranium I, of atomic number 92 and mass number 238, emits an α -particle from its nucleus of charge $+92e$, and hence the new nucleus formed has an atomic number 90 and a mass number 234. This was called uranium X₁, and since the element thorium (Th) has an atomic number 90, uranium X₁ is actually thorium.

A β -particle, an electron, and a γ -ray, an electromagnetic wave, have a negligible effect on the mass of a nucleus when they are emitted. A β -particle has a charge of $-e$. Now uranium X₁ has a nuclear charge of $+90e$ and a mass number 234, and emits β and γ rays. Consequently the mass number is unaltered, but the nuclear charge increases to $+91e$, and hence a new element is formed of atomic number 91. This is uranium X₂ in the series, and is actually the element protactinium. The symbols of the new elements formed are shown in brackets in the column of elements in the Table. The series contains isotopes of uranium (U)₂, lead (Pb), thorium (Th) and bismuth (Bi), that is, elements which have the same atomic numbers but different mass numbers (see p. 1056).

Summarizing, we can say that:

(i) when the nucleus of an element loses an α -particle, the element is displaced two places to the left in the periodic table of the elements, which follows in the order of its atomic number, and lowers its mass number by two units;

(ii) when the nucleus of an element loses a β -particle, the element is displaced one place to the right in the periodic table and its mass number is unaltered.

This law was stated in 1913 by Soddy, Russell and Fajans.

NUCLEAR MASS, NUCLEAR ENERGY

Positive Rays and Atomic Mass

As we saw in the discharge tube (p. 996), at low pressures electrons (negative charges) will flow from the cathode to the anode. If a hole is made in the cathode, some rays appear to pass through the metal, as shown in Fig. 42.18. These rays, which move in the opposite direction

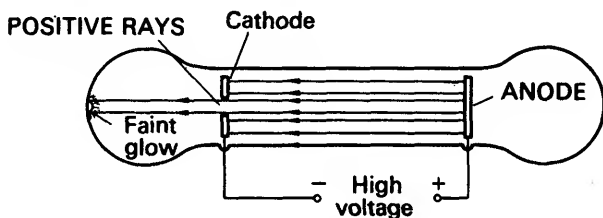


FIG. 42.18. Positive rays.

to the cathode rays (electrons), were called *positive rays*. They were first thought to come from the anode. They are now known to be formed when electrons from the cathode collide with the gas atoms and strip some electrons from the atom. Positive gas ions are then produced. These move slowly towards the cathode on account of the electric field between anode and cathode, and if the cathode is pierced they pass through the hole.

In 1911, Sir. J. J. Thomson measured the masses of individual atoms for the first time. The gas concerned was passed slowly through a bulb at low pressure, and a high voltage was applied. Cathode rays or electrons then flow from the cathode to the anode (not shown), and positive rays due to ionization flow to the cathode, C, whose axis was pierced by a fine tube (Fig. 42.19). The positive rays are *ions*, that is, atoms which have lost one or more electrons. After flowing through C they are subjected to parallel

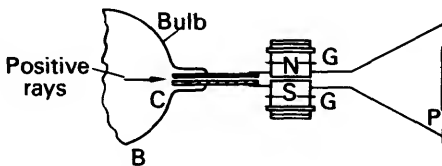


FIG. 42.19. Principle of Thomson's experiment on positive rays.

magnetic and electric fields set at right angles to the incident beam, which are applied between the poles N, S of an electromagnet. Pieces of mica, G, G are used to insulate N, S from the magnet core. The ions were deflected by the fields and were incident on a photographic plate P. After development, parabolic traces were found on P.

Theory

Suppose a positive ray or ion has a charge Q and mass M , and the electric and magnetic field intensities are E , B respectively, acting over a distance D of the ray's path (Fig. 42.20, 42.21).

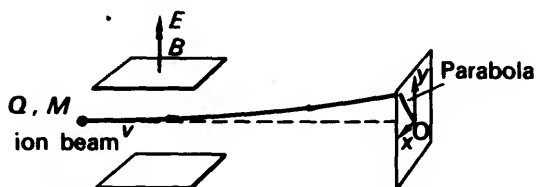


FIG. 42.20. Theory of Thomson's experiment.

(i) From p. 1002, the *electric field* causes the rays to emerge from the plates at an angle θ to the incident direction given by

$$\tan \theta = \frac{QED}{Mv^2},$$

since $e = Q$ here. Hence if y is the deflection in a vertical direction from O, and l is the horizontal distance from the middle of the plates to O,

$$y = l \tan \theta = \frac{QEDl}{Mv^2}. \quad (i)$$

(ii) The *magnetic field* deflects the beam into a circular arc AC of radius R . From p. 1002, this is given by $R = Mv/BQ$.

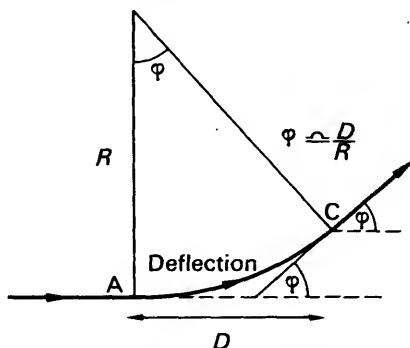


FIG. 42.21. Deflection in magnetic field.

If ϕ is the angle made by the emergent beam with the incident or x -direction, then, if ϕ is small, $\phi = D/R$ (Fig. 42.21). Hence the deflection x in a horizontal direction from O is given by

$$x = l \tan \phi = l\phi = \frac{lD}{R} = \frac{lDBQ}{Mv}. \quad (ii)$$

Eliminating v from (i) and (ii), we obtain

$$x^2 = \frac{Q}{M} \left(\frac{B^2 D l}{E} \right) y \quad (iii)$$

From (iii), it follows that ions with the same charge-mass ratio Q/M , although moving with different velocities, all lie on a *parabola* of the form $x^2 = \text{constant} \times y$.

Determination of Masses. Isotopes

As the zero of the parabola was ill-defined, Thomson reversed the field to obtain a parabolic trace on the other side of the y -axis, as

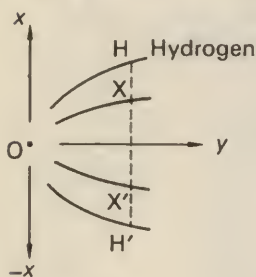


FIG. 42.22. Determination of mass of atom.



FIG. 42.23. Mass Spectrometer. Positive-ray parabolas due to mercury, carbon monoxide, oxygen and carbon ions.

shown in Fig. 42.22. Now from (iii), $x^2 \propto Q/M$ for a given value of y . Consequently any hydrogen ions present would produce the outermost parabola H, since they have the greatest value of Q/M . The masses of ions can thus be measured by comparing the squares of the x -values of the individual parabolas, such as the squares of $X'X$ and $H'H$, for example. In this way Thomson obtained a *mass spectrometer*, one which gave the masses of individual atoms.

With chlorine gas, two parabolas were obtained which gave atomic masses of 35 and 37 respectively. Thus the atoms of chlorine have different masses but the same chemical properties, and these atoms are said to be *isotopes* of chlorine. In chlorine, there are three times as many atoms of mass 35 as there are of mass 37, so that the average atomic weight is $(3 \times 35 + 1 \times 37)/4$, or 35.5. The element xenon has as many as nine isotopes. One part in 5000 of hydrogen consists of an isotope of mass 2 called deuterium, or heavy hydrogen. An unstable isotope of hydrogen of mass 3 is called tritium.

Bainbridge Mass Spectrometer

Thomson's earliest form of mass spectrometer was followed by more sensitive forms. In 1933 Bainbridge devised a mass spectrometer in which the ions were photographed after being deflected by a magnetic field. The principle of the spectrometer is shown in Fig. 42.24.

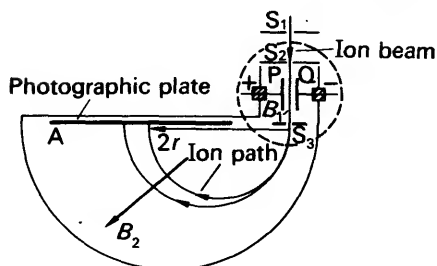


FIG. 42.24. Principle of Bainbridge's mass spectrometer.

Positive ions were produced in a discharge tube (not shown) and admitted as a fine beam through slits S_1 , S_2 . The beam then passed between insulated plates P , Q , connected to a battery, which created an electric field of intensity E . A uniform magnetic field B_1 , perpendicular to E , was also applied over the region of the plates, and all ions, charge \bar{e} , with the same velocity v given by $B_1 \bar{e} v = E \bar{e}$ will then pass

undeflected through the plates and through a slit S_3 . The selected ions are now deflected in a circular path of radius r by a uniform perpendicular magnetic field B_2 , and an image is produced on a photographic plate A , as shown. In this case,

$$\frac{mv^2}{r} = B_2 \bar{e} v.$$

$$\therefore \frac{m}{\bar{e}} = \frac{r B_2}{v}.$$

But for the selected ions, $v = E/B_1$ from above.

$$\therefore \frac{m}{\bar{e}} = \frac{r B_2 B_1}{E}$$

$$\therefore \frac{m}{\bar{e}} \propto r,$$

for given magnetic and electric fields.

Since the ions strike the photographic plate at a distance $2r$ from the middle of the slit S_3 , it follows that the separation of ions carrying the same charge is directly proportional to their mass. Thus a 'linear' mass scale is achieved. A resolution of 1 in 30000 was obtained with a later type of spectrometer.

Einstein's Mass-Energy Relation

In 1905 Einstein showed from his Theory of Relativity that mass and energy can be changed from one form to the other. The energy E produced by a change of mass m is given by the relation:

$$E = mc^2,$$

where c is the numerical value of the velocity of light. E is in joules when m is in kg and c has the numerical value 3×10^8 (p. 560). Thus a change in mass of 1 g could theoretically produce 9×10^{13} joules of energy. Now 1 kilowatt-hour of energy is 1000×3600 or 3.6×10^6 joules, and hence 9×10^{13} joules is 2.5×10^7 or 25 million kilowatt-hours. Consequently a change in mass of 1 g could be sufficient to keep the electric lamps in a million houses burning for about a week in winter, on the basis of about seven hours' use per day.

In electronics and in nuclear energy, the unit of energy called an *electron-volt* (eV) is often used. This is defined as the energy gained by a charge equal to that on an electron moving through a p.d. of one volt.

$$1 \text{ eV} = 1.6 \times 10^{-19} \text{ joule (p. 756).}$$

The *megaelectron-volt* (MeV) is a larger energy unit, and is defined as 1 million eV.

Atomic Mass Unit

If another unit of energy is needed, then one may use a unit of mass, since mass and energy are interchangeable. The *atomic mass unit* (a.m.u.) is defined as one-twelfth of the mass of the carbon atom $^{12}_6\text{C}$. Now the number of molecules in 1 mole of carbon is 6.02×10^{23} , Avogadro's constant, and since carbon is monoatomic, there are 6.02×10^{23} atoms of carbon. These have a mass 12 g.

$$\begin{aligned} \therefore \text{mass of 1 atom of carbon} &= \frac{12}{6.02 \times 10^{23}} \text{ g} = \frac{12}{6.02 \times 10^{26}} \text{ kg} \\ &= 12 \text{ a.m.u.} \end{aligned}$$

$$\begin{aligned} \therefore 1 \text{ a.m.u.} &= \frac{12}{12 \times 6.02 \times 10^{26}} \text{ kg} \\ &= 1.66 \times 10^{-27} \text{ kg.} \end{aligned}$$

We have seen that 1 kg change in mass produces 9×10^{16} joules, and that $1 \text{ MeV} = 1.6 \times 10^{-13}$ joule.

$$\begin{aligned} \therefore 1 \text{ a.m.u.} &= \frac{1.66 \times 10^{-27} \times 9 \times 10^{16}}{1.6 \times 10^{-13}} \text{ MeV} \\ \therefore 1 \text{ a.m.u.} &= 931 \text{ MeV} \end{aligned} \quad (1)$$

This relation is used to change mass units to MeV, and vice-versa, as we shall see shortly. An electron mass, 9.1×10^{-31} kg, corresponds to about 0.5 MeV.

Binding Energy

The protons and neutrons in the nucleus of an atom are called *nucleons*. The work or energy needed to take all the nucleons apart so that they are completely separated is called the *binding energy* of the nucleus. Hence, from Einstein's mass-energy relation, it follows that the total mass of all the individual nucleons is greater than that

of the nucleus, in which they are together. *The difference in mass is a measure of the binding energy.*

As an example, consider a helium nucleus ${}^4_2\text{He}$. This has 4 nucleons, 2 protons and 2 neutrons. The mass of a proton is 1.0076 and the mass of a neutron is 1.009 a.m.u.

$$\therefore \text{total mass of 2 protons plus 2 neutrons} = 2 \times 1.0076 + 2 \times 1.009 \\ = 4.0332 \text{ a.m.u.}$$

But the helium nucleus has a mass of 4.0028 a.m.u.

$$\therefore \text{binding energy} = \text{mass difference of nucleons and nucleus} \\ = 4.0332 - 4.0028 = 0.0304 \text{ a.m.u.} \\ = 0.0304 \times 931 \text{ MeV} = 28.3 \text{ MeV.}$$

The *binding energy per nucleon* of a nucleus is binding energy divided by the total number of nucleons. In the case of the helium nucleus, since there are four nucleons (2 protons and 2 neutrons), the binding energy per nucleon is $28.3/4$ or about 7.1 MeV. Fig. 42.25 shows roughly the variation of the binding energy per nucleon among the elements. The great majority have a value of about 8 MeV per nucleon. In spite of considerable binding energy, elements with high mass numbers may have a tendency to disintegrate. This is not surprising because a very

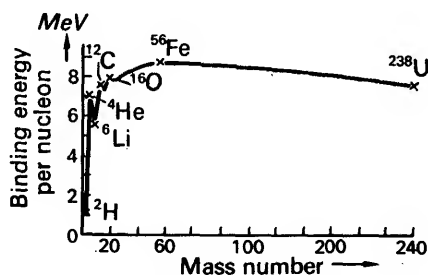
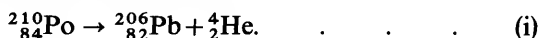


FIG. 42.25. Variation of binding energy per nucleon with mass number.

heavy nucleus contains many protons (and neutrons) packed into a very tiny volume, and strong forces of repulsion may then exist. An α -particle, perhaps formed by two neutron-proton pairs, may then be expelled from the nucleus. A β -particle is emitted when a neutron changes into a proton in the nucleus.

Stability of Nuclei

It is instructive to consider, from an energy point of view, whether a particular nucleus is likely to disintegrate with the mission of an α -particle. As an illustration, consider radium F or polonium, ${}^{210}_{84}\text{Po}$. If an α -particle could be emitted from the nucleus, the reaction products would be the α -particle or helium nucleus, ${}^4_2\text{He}$, and a lead nucleus, ${}^{206}_{82}\text{Pb}$, a reaction which could be represented by:



It should be noted that the sum of the mass numbers, 210, and the

sum of the nuclear charges, $+84e$, of the lead and helium nuclei is equal to the mass number and nuclear charge of the polonium nucleus.

If we require to find whether energy has been released or absorbed in the reaction, we should calculate the total mass of the lead and helium nuclei and compare this with the mass of the polonium nucleus. It is more convenient to use atomic masses rather than nuclear masses, and since the total number of electrons required on each side of (i) to convert the nuclei into atoms is the same, we may use atomic masses in the reaction. These are as follows:

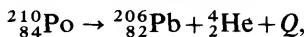
$$\text{lead } {}^{206}_{82}\text{Pb}, = 206.034 \text{ a.m.u.}$$

$$\alpha\text{-particle, } {}^4_2\text{He}, = 4.004 \text{ a.m.u.}$$

$$\therefore \text{total mass} = 210.038 \text{ a.m.u.}$$

Now $\text{polonium } {}^{210}_{84}\text{Po}, = 210.049 \text{ a.m.u.}$

Thus the atomic masses of the products of the reaction are together less than the original polonium nucleus, that is,



where Q is the energy released. It therefore follows that polonium can disintegrate with the emission of an α -particle and a release of energy (see *uranium series*, p. 1053), that is, the polonium is unstable.

Suppose we now consider the possibility of a lead nucleus, ${}^{206}_{82}\text{Pb}$, disintegrating with the emission of an α -particle, ${}^4_2\text{He}$. If this were possible, a mercury nucleus, ${}^{202}_{80}\text{Hg}$, would be formed. The atomic masses are as follows:

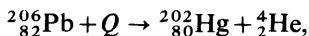
$$\text{mercury, } {}^{202}_{80}\text{Hg}, = 202.035 \text{ a.m.u.}$$

$$\alpha\text{-particle, } {}^4_2\text{He}, = 4.004 \text{ a.m.u.}$$

$$\therefore \text{total mass} = 206.039 \text{ a.m.u.}$$

Now $\text{lead } {}^{206}_{82}\text{Pb}, = 206.034 \text{ a.m.u.}$

Thus, unlike the case previously considered, the atomic masses of the mercury nucleus and α -particle are together *greater* than the lead nucleus, that is,



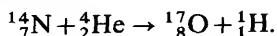
where Q is the energy which must be *given* to the lead nucleus to obtain the reaction products. It follows that the lead nucleus by itself is stable.

Generally, then, a nucleus would tend to be unstable and emit an α -particle if the sum of the atomic masses of the products are together less than that of the nucleus, and it would be stable if the sum of the atomic masses of the possible reaction products are together greater than the atomic mass of the nucleus.

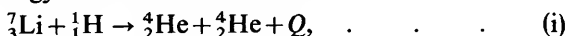
Artificial Disintegration

Uranium, thorium and actinium are elements which disintegrate naturally. The artificial disintegration of elements began in 1919,

when Rutherford used α -particles to bombard nitrogen and found that protons were produced (p. 1051). Some nuclei of nitrogen had changed into nuclei of oxygen, that is, transmutation had occurred, a reaction which can be represented by:



In 1932 Cockcroft and Walton produced nuclear disintegrations by accelerating protons with a high-voltage machine producing about half a million volts, and then bombarding elements with the high-speed protons. When the light element lithium was used, photographs of the reaction taken in the cloud chamber showed that α -particles were produced. The latter shot out in opposite direction from the point of impact of the protons, and as their range in air was equal, the α -particles had initially equal energy. The nuclear reaction was:



where Q is the energy released in the reaction.

To calculate Q , we should calculate the total mass of the lithium and hydrogen nuclei and subtract the total mass of the two helium nuclei. As already explained, however, the total number of electrons required to convert the nuclei to neutral atoms is the same on both sides of equation (i), and hence atomic masses can be used in the calculation in place of nuclear masses. The atomic masses of lithium and hydrogen are 7.018 and 1.008 a.m.u. respectively, a total of 8.026 a.m.u. The atomic mass of the two α -particles is 2×4.004 a.m.u. or 8.008 a.m.u. Thus:

$$\begin{aligned} \text{energy released, } Q, &= 8.026 - 8.008 = 0.018 \text{ a.m.u.} \\ &= 0.018 \times 931 \text{ MeV} = 16.8 \text{ MeV.} \end{aligned}$$

Each α -particle has therefore an initial energy of 8.4 MeV, and this theoretical value agreed closely with the energy of the α -particle measured from its range in air.

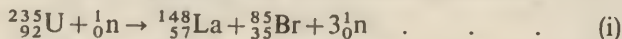
Cockcroft and Walton were the first scientists to use protons for disrupting atomic nuclei after accelerating them by high voltage. Today, giant high-voltage machines are being built at Atomic Energy centres for accelerating protons to enormously high speeds, and the products of the nuclear explosion with light atoms such as hydrogen will yield valuable information on the structure of the nucleus.

Energy released in Fission

In 1934 Fermi began using neutrons to produce nuclear disintegration. These particles are generally more effective than α -particles or protons for this purpose, because they have no charge and are therefore able to penetrate more deeply into the positively-charged nucleus. Usually the atomic nucleus charges only slightly after disintegration, but in 1939 Frisch and Meitner showed that a uranium nucleus had disintegrated into two relatively-heavy nuclei. This is called *nuclear fission*, and as we shall now show, a large amount of energy is released in this case.

Natural uranium consists of about 1 part by weight of uranium

atoms $^{235}_{92}\text{U}$ and 140 parts by weight of uranium atoms $^{238}_{92}\text{U}$. In a nuclear reaction with natural uranium and slow neutrons, it is usually the nucleus $^{235}_{92}\text{U}$ which is fissioned. If the resulting nuclei are lanthanum $^{148}_{57}\text{La}$ and bromine $^{85}_{35}\text{Br}$, together with several neutrons, then:



Now $^{235}_{92}\text{U}$ and ${}_0^1\text{n}$ together have a mass of $(235.1 + 1.009)$ or 236.1 a.m.u. The lanthanum, bromine and neutrons produced together have a mass

$$= 148.0 + 84.9 + 3 \times 1.00 = 235.9 \text{ a.m.u.}$$

∴ energy released = mass difference

$$= 0.2 \text{ a.m.u.} = 0.2 \times 931 \text{ MeV} = 186 \text{ MeV.}$$

$$= 298 \times 10^{-13} \text{ J (approx.).}$$

This is the energy released per atom of uranium fissioned. In 1 kg of uranium there are about

$$\frac{1000}{235} \times 6 \times 10^{23} \text{ or } 26 \times 10^{23} \text{ atoms,}$$

since Avogadro's number, the number of atoms in a mole of any element, is 6.02×10^{23} . Thus if all the atoms in 1 kg of uranium were fissioned, total energy released

$$= 26 \times 10^{23} \times 298 \times 10^{-13} \text{ joules}$$

$$= 2 \times 10^7 \text{ kilowatt-hours (approx.),}$$

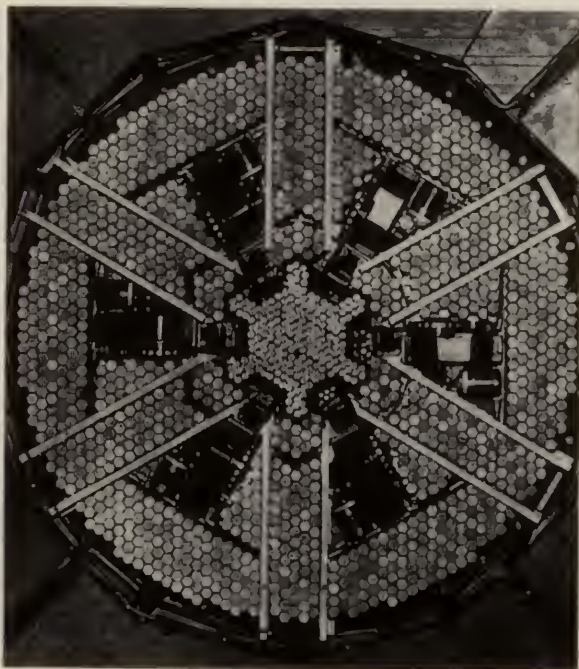


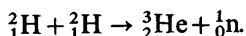
FIG. 42.26. Nuclear Research Reactor, ZEUS. This view shows the heart of the reactor, containing a highly enriched uranium central core surrounded by a natural uranium blanket for breeding studies.

which is the amount of energy given out by burning about 3 million tonnes of coal. The energy released per gramme of uranium fissioned = 8×10^{10} joules (approx.).

To make practical use of nuclear fission, the incident neutrons must be moderated in speed so that they are 'captured' by the nuclei in a mass of uranium. Carbon rods are used as moderators. The neutrons produced in the nuclear reaction in equation (i), p. 1062, in turn produce fission swiftly in other uranium nuclei, and so on, thus creating a multiplying rapid *chain reaction* throughout the mass of uranium. Details of nuclear reactors can be obtained from the United Kingdom Atomic Energy Authority, London.

Energy released in Fusion

In fission, energy is released when a heavy nucleus is split into two lighter nuclei. Energy is also released if light nuclei are *fused* together to form heavier nuclei, and a fusion reaction, as we shall see, is also a possible source of considerable energy. As an illustration, consider the fusion of the nuclei of deuterium, ${}^2_1\text{H}$. Deuterium is an isotope of hydrogen known as 'heavy hydrogen', and its nucleus is called a 'deuteron'. The fusion of two deuterons can result in a helium nucleus, ${}^3_2\text{He}$, as follows:



Now mass of two deuterons

$$= 2 \times 2.015 = 4.03 \text{ a.m.u.},$$

and mass of helium plus neutron

$$= 3.017 + 1.009 = 4.026 \text{ a.m.u.}$$

\therefore mass converted to energy by fusion

$$= 4.03 - 4.026 = 0.004 \text{ a.m.u.}$$

$$= 0.004 \times 931 \text{ MeV} = 3.7 \text{ MeV}$$

$$= 3.7 \times 1.6 \times 10^{-13} \text{ J} = 6.0 \times 10^{-13} \text{ J}$$

\therefore energy released per deuteron

$$= 3.0 \times 10^{-13} \text{ J.}$$

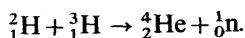
6×10^{23} is the number of atoms in a mole of deuterium, which is about 2 grammes. Thus if all the atoms could undergo fusion,

energy released per gramme

$$= 3.0 \times 10^{-13} \times 3 \times 10^{23} \text{ J}$$

$$= 9 \times 10^{10} \text{ J (approx.).}$$

Other fusion reactions can release much more energy, for example, the fusion of the nuclei of deuterium, ${}^2_1\text{H}$, and tritium, ${}^3_1\text{H}$, isotopes of hydrogen, releases about 30×10^{10} joules of energy according to the reaction:



In addition, the temperature required for this fusion reaction is less

than that needed for the fusion reaction between two deuterons given above, which is an advantage. Hydrogen contains about 1/5000th by weight of deuterium or heavy hydrogen, needed in fusion reactions, and this can be obtained by electrolysis of sea-water, which is cheap and in plentiful supply.

Thermonuclear Reaction

The binding energy curve in Fig. 42.23 shows that elements with low atomic mass, up to about 56, can produce energy by fusion of their nucleons. For fusion to take place, the nuclei must at least overcome their nuclear repulsion when approaching each other. Consequently, for practical purposes, fusion reactions can best be achieved with the lightest elements such as hydrogen, whose nuclei carry the smallest charges and hence repel each other least.

In attempts to obtain fusion, isotopes of hydrogen such as deuterium, ${}^2_1\text{H}$, and tritium, ${}^3_1\text{H}$, are heated to tens of millions of degrees centigrade. The thermal energy of the nuclei at these high temperatures is sufficient for fusion to occur. One technique of promoting this *thermonuclear reaction* is to pass enormously high currents through the gas, which heat it. A very high percentage of the atoms are then ionized and the name *plasma* is given to the gas. Interstellar space or the *aurora borealis* contains a weak form of plasma, but the interior of stars contains a highly concentrated form of plasma. The gas discharge consists of parallel currents, carried by ions, and the powerful magnetic field round one current due to a neighbouring current (see p. 939) draws the discharge together. This is the so-called 'pinch effect'. The plasma, however, wriggles and touches the sides of the containing vessel, thereby losing heat. The main difficulty in thermonuclear experiments in the laboratory is to retain the heat in the gas for a sufficiently long time for a fusion reaction to occur, and the stability of plasma is now the subject of considerable research.

It is believed that the energy of the sun is produced by thermonuclear reactions in the heart of the sun, where the temperature is many millions of degrees centigrade. Bethe has proposed a cycle of nuclear reactions in which, basically, protons are converted to helium by fusion, with the liberation of a considerable amount of energy.

EXERCISES 42

1. $^{24}_{11}\text{Na}$ is a *radioactive isotope* of sodium which has a *half-life period* of 15 hours and disintegrates with the emission of β -particles and γ -rays. It emits β -particles that have energies of 4.2 MeV.)

Explain the meaning of the five terms that are italicized in the statement above. (L.)

2. Give concisely the important facts about mass, charge and velocity associated with α , β and γ radiations respectively. State the effect, if any, of the emission of each of these radiations on (a) the mass number and (b) the atomic number, of the element concerned.

Describe an experiment *either* to measure the range of α particles in air *or* to verify that the intensity of γ radiation varies inversely as the square of the distance from the source, being provided with a suitable radioactive source for the experiment chosen. (L.)

3. 'Gamma rays obey the inverse square law.' What does this mean? What conditions must be satisfied for the statement to be valid?

Describe an experiment designed to verify the statement for gamma rays in air.

The window area of a gamma ray detector is 5.0 cm². The window is placed horizontally and lies 80 cm vertically below a small source of gamma rays, 60 photons per minute from the source are incident on it. Estimate the rate of emission of photons from the source. (N.)

4. Describe the nature of α , β and γ radiations.

What is meant by the statement that the stable isotope of gold has an atomic number of 79 and a mass number of 197? A sample of pure gold is irradiated with neutrons to produce a small proportion of the radioactive isotope of gold of mass number 198. What experiments would you perform to examine the radiation emitted by the sample to establish whether it was α , β or γ radiation? If chemical analysis of the sample subsequently showed that it contained a trace of mercury (atomic number 80) what would you conclude from this about the nature of the radiation from the radioactive gold? What would you expect the mass number of the isotope of mercury present in the gold to be? (O. & C.)

5. Given a standard set of radioactive sources, which include nearly pure α -, β - and γ -emitters, and also a thorium hydroxide preparation, describe with full experimental details how you would demonstrate two of the following:

- (i) that each source emits ionizing radiation;
- (ii) the characteristic differences between α -, β - and γ -radiations;
- (iii) that radioactivity involves a decay process, the half-life of which can be measured;
- (iv) that radioactive decay is a random process. (O.)

6. A Geiger-Müller tube is placed close to a source of beta particles of constant activity. Sketch a graph showing how the count-rate, measured using a suitable scaler or ratemeter, varies with the potential difference applied to the G.M. tube. Discuss how the form of the graph determines the choice of operating conditions for the tube.

Describe how you would investigate the absorption of beta particles of aluminium using a G.M. tube. Sketch a graph showing the results you would expect to obtain. How would the form of the graph change if (a) the same source was used with lead substituted for aluminium, (b) a different source emitting beta particles of higher energy was used, aluminium being the absorbing material? (N.)

7. Compare and contrast the properties of α -particles, protons and neutrons. Discuss briefly the part played by the two last-named particles in atomic structure.

Compare the velocities attained by a proton and an α -particle each of which has been accelerated from rest through the same potential difference. (L.)

8. In nuclear fusion, deuterium nuclei ${}^2_1\text{H}$ might fuse together to form a single helium nucleus. If the atomic masses of deuterium and helium are 2.010 and 4.004 a.m.u. respectively, and 1 a.m.u. = 931 MeV, calculate the energy released in MeV.

9. What is gamma-radiation? Explain one way in which it originates.

An experiment was conducted to investigate the absorption by aluminium of the radiation from a radioactive source by inserting aluminium plates of different thicknesses between the source and a Geiger tube connected to a rate-meter (or scaler). The observations are summarized in the following table:

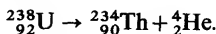
Thickness of aluminium (cm)	Corrected mean count rate (min. ⁻¹)
2.3	1326
6.9	802
11.4	496
16.0	300

Use these data to plot a graph and hence determine for this radiation in aluminium the *linear absorption coefficient*, μ (defined by $\mu = -dI/I \times 1/dx$ where I is the intensity of the incident radiation and dI is the part of the incident radiation absorbed in thickness dx).

Draw a diagram to illustrate the arrangement of the apparatus used in the experiment and describe its preliminary adjustment.

What significance do you attach to the words 'corrected' and 'mean' printed in italics in the table? (N.)

10. Using the information on atomic masses given below, show that a nucleus of uranium 238 can disintegrate with the emission of an alpha particle according to the reaction:



Calculate (a) the total energy released in the disintegration, (b) the kinetic energy of the alpha particle, the nucleus being at rest before disintegration.

Mass of ${}^{238}\text{U}$ = 238.12492 a.m.u. Mass of ${}^{234}\text{Th}$ = 234.11650 a.m.u. Mass of ${}^4\text{He}$ = 4.00387 a.m.u. 1 a.m.u. (atomic mass unit) is equivalent to 930 MeV. (N.)

11. Compare and contrast the properties of the *proton*, *neutron* and *electron*. Explain the role played by each of these particles in the structure of the atom.

How is your account of the arrangement of the electrons in the atom supported by experimental evidence? (L.)

12. The decay of a certain radioactive source is found to be exponential. Explain this statement.

The disintegration of a radioactive atom is considered to be a matter of chance. How do you reconcile this idea with the observed regular law of decay?

To determine the half life of the radioactive gas thoron an experimenter uses an ionization chamber into which he introduces air loaded with thoron. The

ionization current is 90×10^{-12} A initially and falls to half of this value in 60 seconds, which is taken to be the half life of thoron. Unknown to the experimenter, the ionization chamber is contaminated, from previous use, with a radioactive substance whose half life is 11 hours, which is responsible for 4×10^{-12} A of the ionization current at the outset of the experiment. Calculate a more accurate value for the half life of thoron.

What experimental procedure would have revealed this source of error? (O. & C.)

13. What is meant by the *half-life period* (*half-life*) of a radioactive material?

Describe how the nature of α -particles has been established experimentally.

The half-life period of the body polonium-210 is about 140 days. During this period the average number of α -emissions per day from a mass of polonium initially equal to 1 microgram is about 12×10^{12} . Assuming that one emission takes place per atom and that the approximate density of polonium is 10 g cm^{-3} , estimate the number of atoms in 1 cm^3 of polonium. (N.)

14. Describe, with the aid of a labelled diagram, the apparatus used in J. J. Thomson's investigations into the nature and properties of positive rays. Describe, and in qualitative terms explain the pattern observed on the screen of the apparatus.

What conclusions were drawn from the observations made in these experiments? (N.)

15. Give an account of the structure of atoms including the significance of the terms *atomic mass* and *atomic number*. State how the fundamental particles of which atoms are composed differ from each other as regards mass and electric charge.

A potential difference of 600 V is maintained between two identical horizontal metal plates placed 4.0 cm apart one above the other in an evacuated vessel. Particles each with mass $9.1 \times 10^{-31} \text{ kg}$ and electric charge $1.6 \times 10^{-19} \text{ C}$ are emitted with negligible velocity from the plate at the lower potential. For one of the particles calculate (a) the ratio of the electric force to the gravitational force on it, (b) its acceleration, (c) the kinetic energy it acquires on reaching the other plate. (Assume $g = 10 \text{ m s}^{-2}$.) (N.)

chapter forty-three

X-Rays. Photoelectricity. Energy Levels

X-RAYS

X-rays

IN 1895, Röntgen found that some photographic plates, kept carefully wrapped in his laboratory, had become fogged. Instead of merely throwing them aside he set out to find the cause of the fogging. He traced it to a gas-discharge tube, which he was using with a low pressure and high voltage. This tube appeared to emit a radiation that could penetrate paper, wood, glass, rubber, and even aluminium a centimetre and a half thick. Röntgen could not find out whether the radiation was a stream of particles or a train of waves—Newton had the same difficulty with light—and he decided to call it X-rays.

Nature and Production of X-rays

We now regard X-rays as waves, similar to light waves, but of much shorter wavelength: about 10^{-8} cm, or 1 Ångström unit. They are produced when fast electrons, or cathode rays, strike a target, such as the walls or anode of a low-pressure discharge tube. In a modern X-ray tube there is no gas, or as little as high-vacuum technique can achieve: the pressure is about 10^{-5} mm Hg. The electrons are provided by thermionic emission from a white-hot tungsten filament (p. 999). In

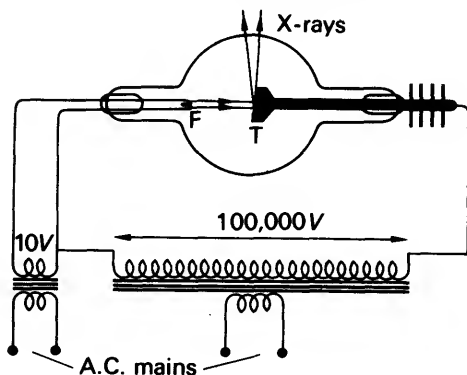


FIG. 43.1. An X-ray tube.

Fig. 43.1 F is the filament and T is the target, or anode. Because there is so little gas, the electrons on their way to the anode do not lose any perceptible amount of their energy in ionizing atoms. From the a.c.

mains, transformers provide about 10 volts for heating the filament, and about 100 000 volts for accelerating the electrons. On the half-cycles when the target is positive, the electrons bombard it, and generate X-rays. On the half-cycles when the target is negative, nothing happens at all—there is too little gas in the tube for it to break down. Thus the tube acts, in effect, as its own rectifier (p. 1010), providing pulses of direct current between target and filament. The heat generated at the target by the electronic bombardment is so great that the target must be cooled artificially. In the figure, fins for air-cooling are shown, but in large tubes the target is made hollow, and is cooled by circulating water or oil. The target in an X-ray tube is usually tungsten, which has a high melting-point.

Effects and Uses of X-rays

When X-rays strike many minerals, such as zinc sulphide, they make them fluoresce. (It was while studying this fluorescence that Becquerel discovered the radiations from uranium.) If a human—or other—body is placed between an X-ray tube and a fluorescent screen, the shadows of its bones can be seen on the screen, because they absorb X-rays more than flesh does. Unusual objects, such as swallowed safety-pins, if they are dense enough, can also be located. X-ray photographs can likewise be taken, with the plate in place of the screen. In this way cracks and flaws can be detected in metal castings.

When X-rays are passed through a crystal, they are scattered by its atoms and diffracted, as light is by a diffraction grating (p. 707). By recording the diffraction pattern on a photographic plate, and measuring it up, the structure of the crystal can be discovered. This was developed by Sir William Bragg and his son, Sir Lawrence Bragg.

X-ray Spectra

In an X-ray tube, very energetic electrons bombard atoms in a metal target such as tungsten, and an electron may be ejected from the innermost shell, the *K* shell. The atom is then in an excited state and is unstable. If an electron from the *L* shell now moves into the vacancy in the *K* shell, the energy of the atom is decreased and simultaneously there is emission of radiation. If E is the change in energy of the atom when the electron moves from the *L* to the *K* shell, then $E = h\nu$, where ν is the frequency of the radiation, from Bohr's theory. Thus $\nu = E/h$, and as E is very high for metals, the frequency ν is very high and the wavelength is correspondingly short. It is commonly of the order of 10^{-8} cm, the wavelengths of X-rays.

The X-ray spectra of different metals such as copper, iron and tungsten are similar in appearance. Each indicates energy changes of electrons in the interior of the atom close to the nucleus. By contrast, the optical spectra of metals are related to the energy changes of electrons in the outermost shells of the atoms, which are different for different metals. The optical spectra are therefore different.

Crystal Diffraction

The first proof of the wave-nature of X-rays was due to Laue in 1913, many years after X-rays were discovered. He suggested that the regular small spacing of atoms in crystals might provide a natural diffraction grating if the wavelengths of the rays were too short to be used with an optical line grating. Experiments by Friedrich and Knipping showed that X-rays were indeed diffracted by a thin crystal, and produced a pattern of intense spots round a central image on a photographic plate placed to receive them (Fig. 43.2). The rays had thus

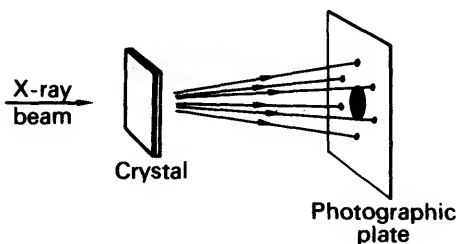


FIG. 43.2. Laue crystal diffraction.

been scattered by interaction with electrons in the atoms of the crystal, and the pattern obtained gave information on the geometrical spacing of the atoms.

Bragg's Law

The study of the atomic structure of crystals by X-ray analysis was initiated in 1914 by Sir William Bragg and his son Sir Lawrence Bragg, with notable achievements. They soon found that a monochromatic beam of X-rays was reflected from a plane in the crystal rich in atoms, as if the latter acted like a mirror.

This important effect can be explained by Huyghens's wave theory in the same way as the reflection of light by a plane surface. Suppose a monochromatic parallel X-ray beam is incident on a crystal and interacts with atoms such as A, B, C, D in an atomic plane P (Fig. 43.3 (i)). Each atom scatters the X-rays. Using Huyghens's construction, wavelets can be drawn with the atoms as centres, which all lie on a plane

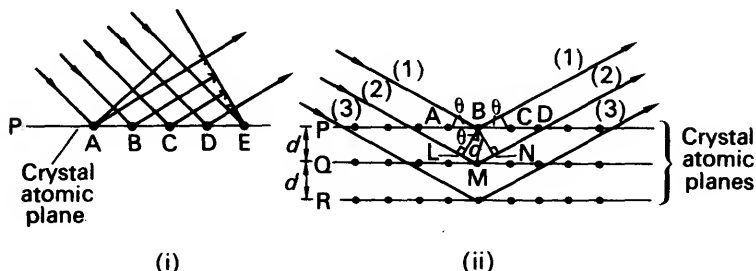


FIG. 43.3. Reflection (diffraction) at crystal atomic planes.

wavefront reflected at an equal angle to the atomic plane P. When the X-ray beam penetrates the crystal to other atomic planes such as Q, R parallel to P, reflection occurs in a similar way (Fig. 43.3 (ii)). Usually, the beam or ray reflected from one plane is weak in intensity. If, however, the reflected beams or rays from all planes are *in phase* with each other, an intense reflected beam is produced by the crystal.

Suppose, then, that the glancing angle on an atomic plane in the crystal is θ , and d is the distance apart of consecutive parallel atomic planes (Fig. 43.3 (ii)). The path difference between the rays marked (1) and (2) = $LM + MN = 2LM = 2d \sin \theta$. Thus an intense X-ray beam is reflected when

$$2d \sin \theta = n\lambda,$$

where λ is the wavelength and n has integral values. This is known as *Bragg's law*. Hence, as the crystal is rotated so that the glancing angle is increased from zero, and the beam reflected at an equal angle is observed each time, an intense beam is suddenly produced for a glancing angle θ_1 such that $2d \sin \theta_1 = \lambda$. When the crystal is rotated further, an intense reflected beam is next obtained for an angle θ_2 when $2d \sin \theta_2 = 2\lambda$. Thus several orders of diffraction images may be observed. Many orders are obtained if λ is small compared with $2d$. Conversely, no images are obtained if λ is greater than $2d$.

The intense diffraction (reflection) images from an X-ray tube are due to X-ray lines *characteristic of the metal used as the target*, or 'anti-cathode' as it was originally known. This is because the quantum of energy $h\nu$ in the emitted X-ray depends on the nuclear charge $+Ze$ of the atom, which affects electron energy changes near the nucleus (p. 1084). The frequency, or wavelength λ , thus depends on the atomic number, Z .

X-ray Analysis

A special form of spectrometer was designed by Sir William Bragg for his experiments. The crystal was fixed on the table, and an X-ray beam, limited by lead shields A, B, was incident on the crystal at various glancing angles, θ (Fig. 43.4 (i)). An ionization chamber, Q, was used to measure the intensity of the X-rays reflected by the crystal. Q contained a heavy gas such as methyl iodide, and the intensity of the

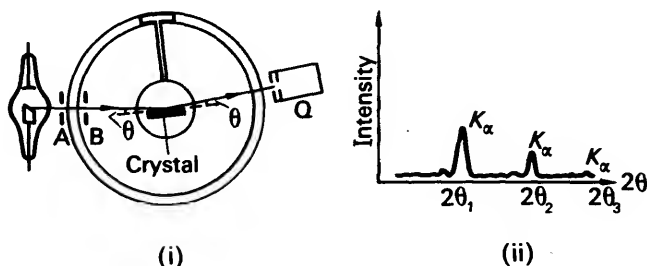


FIG. 43.4. X-ray spectrometer and results.

X-rays was proportional to the ionization current flowing, which was measured by means of an electrometer (not shown) connected to Q. The table and Q were geared so that Q turned through twice the angle of rotation of the crystal, and was always ready to measure the intensity of X-rays satisfying the law $2d \sin \theta = n\lambda$.

Typical results with particular parallel atomic planes in a crystal such as sylvine (KCl) or rocksalt (NaCl) are shown roughly in Fig. 43.4 (ii). A characteristic X-ray line such as K_α produces peaks of intensity at glancing angles θ_1 , θ_2 and θ_3 for the first three orders. Measurement shows that $\sin \theta_1 : \sin \theta_2 : \sin \theta_3 = 1 : 2 : 3$, thus verifying Bragg's law, $2d \sin \theta = n\lambda$.

Crystal Atomic Spacing

Before the wavelength λ can be calculated, the distance d between consecutive parallel atomic planes is required. As an illustration of the calculation, consider the distance d between those atomic planes of a rock-salt crystal, NaCl, which are parallel to the face ABCD of a unit cell or cube of the crystal (Fig. 43.5). In this case $d = a$, the side of the cube. We thus require the distance a between consecutive atoms (ions) of sodium and chlorine in the crystal.

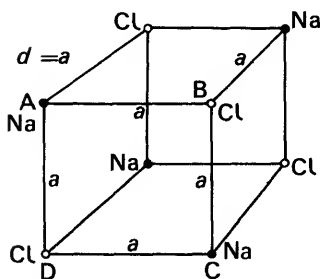


FIG. 43.5. Unit cell of rock salt.

The mass of one mole of sodium chloride is 58.5 g, the sum of the relative atomic masses of sodium and chlorine. This contains about 6×10^{23} molecules, Avogadro's constant. The mass of a molecule is thus $58.5 \text{ g} / 6 \times 10^{23}$. Since the density of rocksalt is about 2.2 g cm^{-3} ,

$$\text{volume occupied by 1 molecule (2 atoms)} = \frac{58.5}{6 \times 10^{23} \times 2.2} \text{ cm}^3.$$

$$\therefore \text{volume associated with each atom} = \frac{58.5}{2 \times 6 \times 10^{23} \times 2.2} \text{ cm}^3.$$

$$\therefore \text{separation of atoms} = \left[\frac{58.5}{2 \times 6 \times 10^{23} \times 2.2} \right]^{1/3} \text{ cm.}$$

$$= 2.8 \times 10^{-8} \text{ cm.}$$

Thus if the first order diffraction image is obtained for a glancing angle θ of 5.4° for a particular X-ray wavelength λ , then

$$\lambda = 2d \sin \theta = 2 \times 2.8 \times 10^{-8} \times \sin 5.4^\circ$$

$$= 0.5 \times 10^{-8} \text{ cm} = 0.5 \text{ \AA}.$$

Knowing λ , the atomic spacing d in other crystals can then be found, thus leading to analysis of crystal structure.

Moseley's Law

In 1914 Moseley measured the frequency ν of the characteristic X-rays from many metals, and found that, for a particular type of emitted X-ray such as K_{α} , the frequency ν varied in a regular way with the atomic number Z of the metal. When a graph of Z v. $\nu^{1/2}$ was plotted, an almost perfect straight line was obtained (Fig. 43.6). Moseley therefore gave an empirical relation, known as *Moseley's law*, between ν and Z as

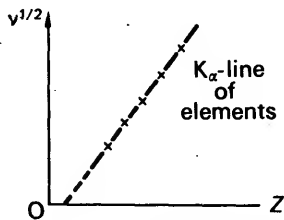


FIG. 43.6. Moseley's law.

$$\nu = a(Z - b)^2,$$

where a , b are constants.

Since the regularity of the graph was so marked, Moseley predicted the discovery of elements with atomic numbers 43, 61, 72 and 75, which were missing from the graph at that time. These were later discovered. He also found that though the atomic weights of iron, nickel and cobalt increased in this order, their positions from the graph were: iron ($Z = 26$), cobalt ($Z = 27$) and nickel ($Z = 28$). The chemical properties of the three elements agree with the order by atomic number and not by atomic weight. Rutherford's experiments on the scattering of α -particles (p. 1050) showed that the atom contained a central nucleus of charge $+Ze$ where Z is the atomic number, and Moseley's experiments confirm the importance of Z in atomic theory (see also p. 1090).

Continuous X-ray Background Radiation

The characteristic X-ray spectrum from a metal is usually superimposed on a background of continuous, or so-called 'white', radiation of small intensity. Fig. 43.7 illustrates the characteristic lines, K_{α} , K_{β} , of a metal and the continuous background of radiation for two values of p.d., 40000 and 32000 volts, across an X-ray tube. It should be

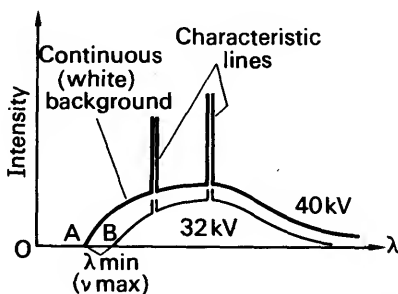


FIG. 43.7. X-ray characteristic lines and background.

noted that (i) the wavelengths of the characteristic lines are independent of the p.d.—they are characteristic of the metal, (ii) the background of

by Davisson and Germer, show that streams of *electrons* produce diffraction patterns and hence also exhibit wave properties. Electron diffraction is now as useful a research tool as X-ray diffraction.

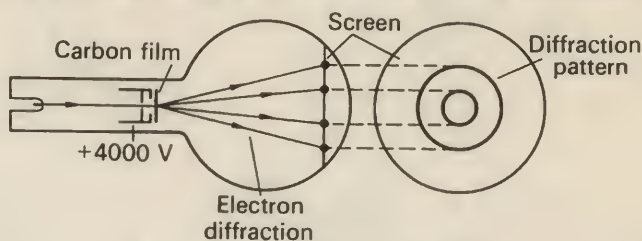
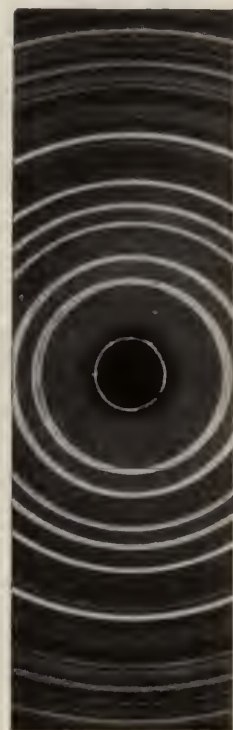


FIG. 43.8 (a). Electron diffraction tube.

A TELTRON tube available for demonstrating electron diffraction, is shown diagrammatically in Fig. 43.8 (a). A beam of electrons impinges on a layer of graphite which is extremely thin, and a diffraction pattern, consisting of rings, is seen on the tube face. Sir George Thomson first obtained such a diffraction pattern using a very thin gold film. If the voltage V on the anode is increased, the velocity, v , of the electrons is increased. The rings are then seen to become narrow, showing that the wavelength λ of the electron waves decreases with increasing v or increasing voltage V .



(i) X-ray diffraction rings produced by a crystal.

FIG. 43.8 (b).
Similarity of
Wave (i)
and
Particle (ii)



(ii) Electron diffraction rings produced by a thin gold film.

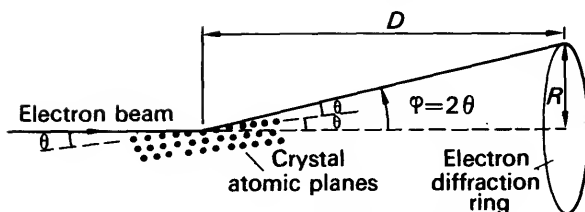


FIG. 43.9. Theory of diffraction experiment.

If a particular ring of radius R is chosen, the angle of deviation ϕ of the incident beam is given by $\phi = 2\theta$, where θ is the angle between the incident beam and the crystal planes. Fig. 43.9. Now $\tan \phi = R/D$, and if ϕ is small, $\phi = R/D$ to a good approximation. Hence $\theta = R/2D$. If the Bragg law is true for electron diffraction as well as X-ray diffraction, then, with the usual notation, $2d \sin \theta = n\lambda$.

$$\therefore \lambda \propto \sin \theta \propto \theta \propto R. \quad (i)$$

On plotting a graph of R against $1/\sqrt{V}$ for different values of accelerating voltage V , a straight line graph passing through the origin is obtained. Now $\frac{1}{2}m_e v^2 = eV$, or $1/\sqrt{V} \propto 1/v$, where v is the velocity of the electrons accelerated from rest. Hence the electrons appear to act as waves whose wavelength is inversely-proportional to their velocity. This is in agreement with de Broglie's theory, now discussed.

De Broglie's Theory

In 1925, before the discovery of electron diffraction, de Broglie proposed that

$$\lambda = \frac{h}{p}, \quad (ii)$$

where λ is the wavelength of waves associated with particles of momentum p , and h is Planck's constant, 6.63×10^{-34} joule second. The quantity h was first used by Planck in his theory of heat radiation and is a constant which enters into all branches of atomic physics. It is easy to see that de Broglie's relation is consistent with the experimental result (ii). The gain in kinetic energy is eV so that

$$\frac{1}{2}m_e v^2 = eV,$$

where v is the velocity of the electrons. Thus $v = \sqrt{2eV/m_e}$ and hence

$$p = m_e v = \sqrt{2eVm_e}$$

$$\therefore \lambda = \frac{h}{p} = \frac{h}{m_e v} = \frac{h}{\sqrt{2eVm_e}} \propto V^{-1/2}.$$

We can now estimate the wavelength of an electron beam. Suppose $V = 3600$ volts. For an electron, $m = 9.1 \times 10^{-31}$ kg, $e = 1.6 \times 10^{-19}$ coulomb, and $h = 6.6 \times 10^{-34}$ joule second.

$$\begin{aligned} \therefore \lambda &= \frac{h}{\sqrt{2eVm_e}} = \frac{6.6 \times 10^{-34}}{\sqrt{2 \times 1.6 \times 10^{-19} \times 3600 \times 9.1 \times 10^{-31}}} \\ &= 2 \times 10^{-11} \text{ metres.} \end{aligned}$$

This is about 30000 times smaller than the wavelength of visible light. On this account electron beams are used in *electron microscopes*. These instruments can produce resolving powers far greater than that of an optical microscope.

Wave Nature of Matter

Electrons are not the only particles which behave as waves. The effects are less noticeable with more massive particles because their momenta are generally much higher, and so the wavelength is correspondingly shorter. Since appreciable diffraction is observed only when the wavelength is of the same order as the grating spacing, the heavier particles, such as protons, are diffracted much less. Slow neutrons, however, are used in diffraction experiments, since the low velocity and high mass combine to give a momentum similar to that of electrons used in electron diffraction. The wave nature of α -particles is important in explaining α -decay.

PARTICLE NATURE OF WAVES

We have already seen that γ -rays behave as electromagnetic waves of very short wavelength. Now γ -rays can be detected by Geiger-Müller (GM) tubes and solid state detectors, where *individual* pulses are counted. Thus, on detection, γ -rays behave as *particles*. Other evidence for the particle nature of electromagnetic waves is given by the photoelectric effect, which we now discuss.

Photoelectricity

In 1888 Hallwachs discovered that an insulated zinc plate, negatively charged, lost its charge if exposed to ultra-violet light. Hertz had previously noticed that a spark passed more easily across the gap of an induction coil when the negative metal terminal was exposed to sunlight. Later investigators such as Lenard and others showed that electrons were ejected from a zinc plate when exposed to ultra-violet light. Light thus gives energy to the electrons in the surface atoms of the metal, and enables them to break through the surface. This is called the *photoelectric effect*.

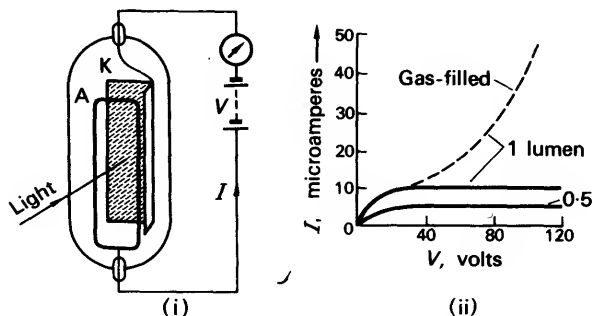


FIG. 43.10. (i) A photo-cell (photo-emissive). (ii) Characteristic.

Fig. 43.10 (i) is a diagram of a modern photoelectric cell. The cathode K is a V-shaped plate coated with caesium or some more complicated and very sensitive surface. In front of it is a wire ring, the anode A, which collects the photo-electrons. The heavy curve at (ii) in the figure shows how the current through the cell, I , varies with the potential difference across it, V . At first the current rises, but at a potential difference of about 30 volts it saturates. We suppose that the anode is then collecting all the electrons emitted by the cathode. This curve is drawn for a light-flux of 1 lumen upon the cathode (p. 562). If the flux is halved, the saturation current also falls by a half, showing that the number of electrons emitted per second is proportional to the light-flux falling upon the cathode.

Some photoelectric cells contain an inert gas—such as argon—at a pressure of a few millimetres of mercury. The current in such a cell does not saturate, because the electrons ionize the gas atoms by collision. Fig. 43.10 (ii). The greater the potential difference, the greater the kinetic energy of the electrons, and the more intense the ionization of the gas.

Photoelectric cells are used in photometry, in industrial control and counting operations, in television, and in many other ways. Their use in reproducing sound from film is explained in the Sound section of this book (p. 599).

Photo-voltaic Cells

Photoelectric cells of the kind we have just described are called photo-emissive cells, because in them light causes electrons to be emitted. Another type of cell is called photo-voltaic, because it generates an e.m.f. and can therefore provide a current without a battery. One form of such a cell consists of a copper disc, oxidized on one

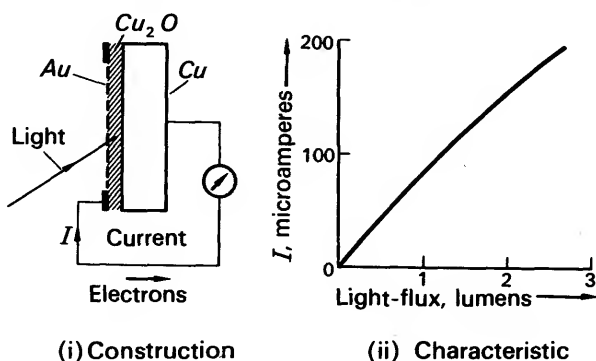


FIG. 43.11. A photo-voltaic cell.

face ($\text{Cu}_2\text{O}/\text{Cu}$), as shown in Fig. 43.11 (i). Over the exposed surface of the oxide a film of gold (Au) is deposited, by evaporation in a vacuum; the film is so thin that light can pass through it. When it does so it generates an e.m.f. in a way which we cannot describe here.

Photo-voltaic cells are sensitive to visible light. Fig. 43.11 (ii) shows

how the current from such a cell, through a galvanometer of resistance about 100 ohms, varies with the light-flux falling upon it. The current is not quite proportional to the flux. Photo-voltaic cells are obviously convenient for photographic exposure meters, for measuring illumination in factories, and so on, but as measuring instruments they are less accurate than photo-emissive cells.

Photo-conductive Cells

A photo-conductive cell is one whose resistance changes when it is illuminated. A common form consists of a pair of interlocking comb-like electrodes made of gold (Au) deposited on glass (Fig. 43.12); over these a thin film of selenium (Se) is deposited. In effect, the selenium

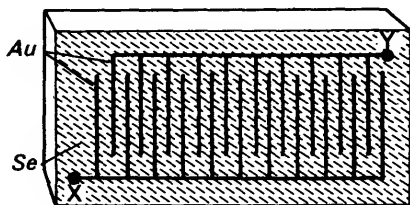


FIG. 43.12. A selenium cell.

forms a large number of strips, electrically in parallel; this construction is necessary because selenium has a very high resistivity (about 700 ohm m in the dark). The resistance between the terminals, XY, falls from about 10^7 ohms in the dark to about 10^6 ohms in bright light. In conjunction with valve amplifiers, photo-conductive cells were used as fire alarms during the last war. They were the first photo-cells to be discovered—in 1873—but they were the least useful. They are sluggish, taking about a second to respond fully to a change of illumination; and they show hysteresis—one change of illumination affects their response to the next.

Velocity of Photo-electrons

In 1902 Lenard found that the velocity of ejection of the electron from an illuminated metal was independent of the intensity of the particular incident monochromatic light. It appeared to vary only with the wavelength or frequency of the incident light, and above a particular wavelength no electrons were emitted. This was a very surprising result. It could not be explained on classical grounds, which predicts that the greater the light energy incident on the metal, the greater should be the energy of the liberated electrons, and that electrons should always be ejected, irrespective of the incident wavelength, if the incident energy is large enough.

Theories of Light. The Photon

About 1660 Newton had proposed a *corpuscular theory* of light, that is, light consists of particles or corpuscles, and he explained the phenomena of reflection and refraction by applying the laws of mech-

anics to the particles (p. 680). About the same period Huyghens proposed a *wave theory* of light, that is, light travels by the propagation of a wave or disturbance in the medium (p. 676), and this was applied with particular success to the phenomena of interference and diffraction. Newton's theory was abandoned soon after 1800 when Thomas Young revived interest in Huyghen's wave theory. Among other difficulties, Newton's theory led to the conclusion that the velocity of light in water was greater than in air, which was shown to be untrue experimentally.

In 1902 Planck had shown that the experimental observations in black-body radiation could be explained on the basis that the energy from the body was emitted in separate or discrete packets of energy, known as *quanta* of energy, of amounts $h\nu$, where ν is the frequency of the radiation and h is a constant known as *Planck's constant*. This is the *quantum theory of radiation*. With characteristic genius, Einstein asserted in 1905 that the unexpected experimental result of Lenard—that the energy of the ejected electron was independent of the intensity of the incident light and depended only on the frequency of the light—could be explained by applying a quantum theory of light. He assumed that light of frequency ν contains packets or quanta of energy $h\nu$. On this basis, light consists of *particles*, and these are called *photons*. The number of photons per unit area of cross-section of the beam of light per unit time is proportional to its intensity, but the energy of a photon is proportional to its frequency.

The minimum amount of work or energy to take a free electron out of the surface of a metal against the attractive forces of the positive ions is known as the *work function*, w_0 , of the metal. When light of sufficiently high frequency is incident on the metal, an amount w_0 of the incident energy $h\nu$ is used to liberate the electron, leaving an excess energy $h\nu - w_0$, which is given to the ejected electron. The maximum kinetic energy, $\frac{1}{2}m_e v_{\max}^2$, of the latter is thus, on Einstein's theory:

$$\frac{1}{2}m_e v_{\max}^2 = h\nu - w_0. \quad (i)$$

Millikan's Experiment

To test the linear relationship between the kinetic energy of the ejected electron and the frequency expressed in (i), Millikan carried out experiments in 1916 using the alkali metals lithium, sodium and potassium. These metals emit electrons when illuminated by ordinary (visible) light, and cylinders of them, A, B, C, were placed round a wheel W (Fig. 43.13). To avoid tarnishing and the formation of

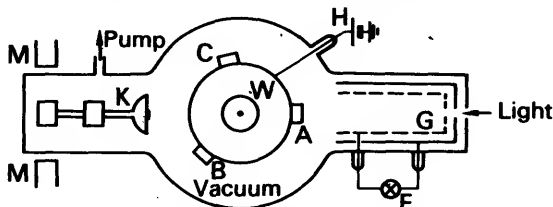


FIG. 43.13. Millikan's photoelectric experiment.

oxide films on the metal surface, which lead to considerable error, the metals were housed in a vacuum. Their surfaces were kept clean by a cutting knife K, which could be moved and turned by means of a magnet M outside.

The metal, A say, was kept at a variable positive potential by a battery H, and illuminated by a beam of monochromatic light of wavelength λ_1 from a spectrometer. Any photo-electrons emitted could reach a gauze cylinder G, which was connected to one side of an electrometer E whose other terminal was earthed, and a current I would then flow in E. When the potential of A is increased, G has an increasing negative potential relative to A, $-V$ say, and the current I then decreases. At some negative value, $-V_1$, the current becomes zero (Fig. 43.14 (i)). The potential of G is then the same as the other terminal of E, which is earthed, and the negative potential of G relative to A is thus now given numerically by the p.d. of the battery H. Millikan obtained variations of current, I , with monochromatic light of other wavelengths λ_2, λ_3 , using light of constant intensity.

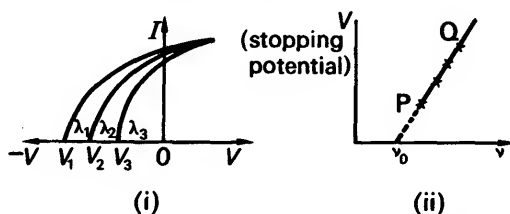


FIG. 43.14. Results of Millikan's experiment.

Deduction from Millikan's Results

The negative potential V of G relative to A when no electrons reach G is called the 'stopping potential' of G. In this case the maximum kinetic energy of the ejected electrons is just equal to the work eV done in moving against the opposing p.d. Thus

$$eV = h\nu - w_0. \quad (i)$$

This is a linear relation between V and ν , and when the stopping potential was plotted against the frequency, a straight line PQ was obtained (Fig. 43.14 (ii)). Now from (i), the slope of the line is h/e , and knowing e , Millikan calculated h . The result was 6.26×10^{-34} joule second, which was very close to the value of h found from experiments on black-body radiation. This confirmed Einstein's photoelectric theory that light can be considered to consist of particles with energy $h\nu$.

In (i), we can write the work function energy w_0 as $h\nu_0$, where $\nu_0 = w_0/h$. Hence, for the electrons with maximum energy,

$$eV = \text{kinetic energy of electron} = h\nu - w_0 = h(\nu - \nu_0).$$

It then follows that no electrons are emitted from a metal when the incident light has a frequency less than ν_0 . The magnitude of ν_0 is called the *threshold frequency* of the metal concerned, and is given by the intercept of PQ with the axis of ν (Fig. 43.14 (ii)).

EXAMPLE

Caesium has a work function of 1.9 electron-volts. Find (i) its threshold wavelength, (ii) the maximum energy of the liberated electrons when the metal is illuminated by light of wavelength 4.5×10^{-5} cm (1 electron-volt = 1.6×10^{-19} J, $h = 6.6 \times 10^{-34}$ J s, $c = 3.0 \times 10^8$ m s $^{-1}$).

(i) The threshold frequency, ν_0 , is given by $h\nu_0 = w_0 = 1.9 \times 1.6 \times 10^{-19}$ J. Now threshold wavelength, λ_0 , = c/ν_0

$$\begin{aligned}\therefore \lambda_0 &= \frac{c}{w_0/h} = \frac{ch}{w_0} \\ &= \frac{3 \times 10^8 \times 6.6 \times 10^{-34}}{1.9 \times 1.6 \times 10^{-19}} \\ &= 6.5 \times 10^{-7} \text{ m.}\end{aligned}$$

(ii) Maximum energy of liberated electrons = $h\nu - w_0$, where ν is the frequency of the incident light. But $\nu = c/\lambda$.

$$\begin{aligned}\therefore \text{max. energy} &= \frac{hc}{\lambda} - w_0 \\ &= \frac{6.6 \times 10^{-34} \times 3 \times 10^8}{4.5 \times 10^{-7}} - 1.9 \times 1.6 \times 10^{-19} \\ &= 1.4 \times 10^{-19} \text{ J.}\end{aligned}$$

Duality

From what has been said, it is clear that particles can exhibit wave properties, and that waves can sometimes behave as particles. It would appear, therefore, that a paradox exists since wave and particle structure appear mutually exclusive.

Scientists gradually realized, however, that the dual aspect of wave-particle properties are completely general in nature. All physical entities can be described either as waves or particles; the description to choose is entirely a matter of convenience. The two aspects, wave and particle, are linked through the two relations

$$E = h\nu; \quad p = h/\lambda.$$

On the left of each of these relations, E and p refer to a particle description. On the right, ν and λ refer to a wave description. Note that Planck's constant is the constant of proportionality in *both* these equations, a fact which can be predicted by Einstein's Special Theory of Relativity.

QUANTIZATION OF ENERGY

Energy of Atoms

The average energy of a monatomic molecule moving in a gas at room temperature is $\frac{3}{2}kT$ or $\frac{3}{2} \times 1.4 \times 10^{-23} \times 300$ joule, which is 6.3×10^{-21} joule. Since 1 eV is 1.6×10^{-19} joule, this energy corresponds to about 0.04 eV. Thus when collisions between molecules take place, the energy exchange is of the order of 0.04 eV. In these conditions the collisions are *perfectly elastic*, that is, the internal energy of the

atoms is not increased and all the energy remains in the form of translational kinetic energy.

In 1914, Franck and Hertz bombarded atoms by electrons of much higher energy, of the order of several electron volts. They used sodium vapour at a very low pressure of about 1 mm of mercury in a tube containing a heated tungsten filament F, a grid plate G, and a plate A (Fig. 43.15 (i)). Electrons were emitted from F, and the distance FG was arranged to be much greater than the mean free path of the electrons in the gas, in which case the electrons would make collisions with the atoms before reaching G. The p.d. V between F and G could be varied by the potentiometer S. The electrons emitted from F were accelerated to kinetic energies depending on the magnitude of V , measured by a voltmeter. A small p.d., less than 1 volt, was applied

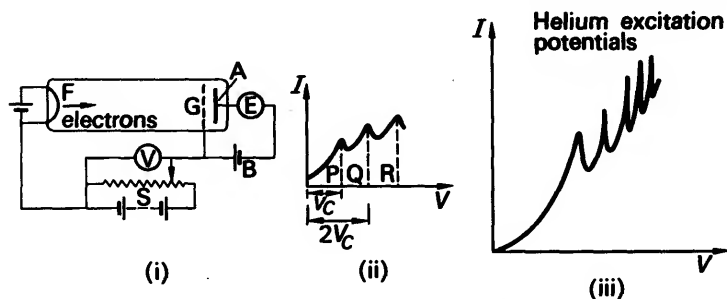


FIG. 43.15. Franck and Hertz experiment.

between A and G so that A was negative in potential relative to G. The plate A was close to G, and electrons reaching G and passing through to A were subjected to a retarding field. The number per second reaching A was measured by an electrometer E.

When the accelerating p.d. V between F and G was increased from zero, the current in E rose until the p.d. reached a value P (Fig. 43.15 (ii)). As V was increased further the current diminished to a minimum, rose again to a new peak at a higher p.d. Q, then diminished again and rose to another peak at a higher p.d. R. The p.d. V_c between successive peaks was found to be constant and equal to 2.10 volts for sodium vapour. Similar results were found for other gases.

Energy Levels

From the graph, it can be seen that the current begins to drop at the critical potential V_c . Here the electrons have an energy of V_c electron-volts or eV_c joules. This energy is just sufficient to raise the internal energy of the sodium atom by collision. Energies less than eV_c fail to increase the energy of the atom. After giving up this energy, the electrons then have insufficient energy to overcome the small retarding p.d. between G and A. Thus the current starts to fall. At a p.d. of $2V_c$ a dip again begins to form. This is due to the electrons giving up energy equal to $2eV_c$ to two atoms at separate collisions.

It thus appears that the energy of the atom cannot be increased

unless the energy of the colliding particle is greater than V_c electron-volts. In this case an *inelastic* collision takes place. The atom now takes up an energy equal to eV_c joules, which the colliding particle loses. This process of increasing the energy of an atom is called *excitation*. The interval, V_c between successive peaks of the graph is the *excitation potential* of the atom.

The results of the Franck-Hertz experiment show that the energy of an atom is constant unless the atom is given enough energy to raise this by a *definite* amount. *No intermediate energy change is allowed*. An atom, therefore, exists in one of a set of well defined *energy levels*. If helium, for example, is used in a Franck-Hertz tube, a graph shown in Fig. 43.15 (iii) is obtained. Here each peak corresponds to a different energy level of the atom. Thus a whole sequence of different energy levels can be found in the helium atom.

As we have seen, at ordinary temperatures, the thermal energy of molecules in a gas is insufficient to cause excitation. If the gas can be heated to an enormously high temperature, of the order of 100 000 K. the molecules can gain enough energy to cause excitation.

Energy Levels in Spectra

If a gas is excited by a high voltage to produce a discharge, and the light is examined in a spectrometer, an emission spectrum is seen. A number of gases such as neon produce a *line spectrum*, that is, the spectrum consists of a number of well defined lines, each having a particular wavelength or frequency. These lines are also experimental evidence for the existence of separate or 'quantized' energy levels in the atom, as we now explain.

As they move through the discharge in the gas, some electrons have sufficient energy to excite atoms to a higher energy level. We suppose that a given atom has a series of defined and discrete (separated) energy levels of the atom, E_0, E_1, E_2 , and that no other or intermediate energy level is possible. The lowest energy level E_0 is called the *ground state energy*. All physical systems are in stable equilibrium in the lowest energy state. Thus once an atom has been excited to a higher energy level E_n , it will try to reduce its energy. The energy lost if the atom reverts directly to the ground state is $(E_n - E_0)$. *This energy is radiated in the form of a photon of electromagnetic radiation*. The energy of the photon is $h\nu$ where ν is the frequency of the radiation, and thus

$$h\nu = E_n - E_0.$$

It can now be seen that a number of frequencies ν of electromagnetic radiation may be produced from a hot gas in a discharge tube. Each frequency corresponds to a possible energy change in the atom. Sometimes it is possible for the energy to change back to the ground state *via* an intermediate energy level E_m . In this case two different frequencies ν_1, ν_2 are radiated which are given respectively by the equations:

$$h\nu_1 = E_n - E_m, \quad h\nu_2 = E_m - E_0.$$

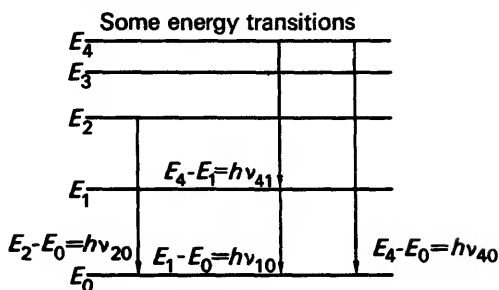


FIG. 43.16. Energy transitions.

It is customary to draw the energy levels on a vertical scale and to mark the *transitions* from one energy level to another with an arrow. Fig. 43.16.

Bohr's Theory of Hydrogen Atom

A *model* of the hydrogen atom was proposed by Bohr in 1911. This explained satisfactorily the existence of energy levels and the spectrum of the hydrogen atom. Later, however, it was shown that the model could not be applied to other atoms and a more satisfactory 'quantum theory' of the atom has since been developed.

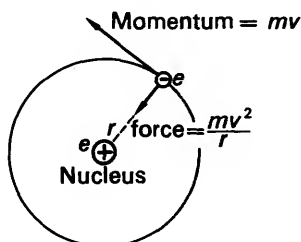


FIG. 43.17. Bohr's theory of hydrogen atom.

Bohr considered one electron of charge $-e$ and mass m , moving in a circular orbit round a central hydrogen nucleus of charge $+e$ (Fig. 43.17). The energy of the electron is partly kinetic and partly potential. If v is the velocity in the orbit, then

$$\text{kinetic energy} = \frac{1}{2} mv^2 \quad . \quad . \quad . \quad (i)$$

If the electron is removed a very long way against the attraction of the nucleus, that is, to infinity, its potential energy is a maximum, but in calculations this energy is given the value of 'zero'. In practice, this means that the hydrogen atom then loses an electron completely and becomes an *ion*. Consider now an electron in the atom at a distance r from the nucleus. Since work is required to move the electron from this point to infinity against the attraction of the nucleus, it follows that the potential energy of the electron is *negative*. The potential due

to the nuclear charge $+e$ at a distance r is given by $+e/4\pi\epsilon_0 r$. Since this is the work done per unit charge, then

$$\text{potential energy of electron} = \frac{e}{4\pi\epsilon_0 r} \times -e = \frac{-e^2}{4\pi\epsilon_0 r} \quad (\text{ii})$$

Hence, from (i) and (ii),

$$\text{total energy, } E = \frac{1}{2}mv^2 - \frac{e^2}{4\pi\epsilon_0 r} \quad (\text{iii})$$

For circular motion,

$$\frac{mv^2}{r} = \text{centripetal force} = \frac{e^2}{4\pi\epsilon_0 r^2} \quad (\text{iv})$$

$$\therefore \frac{1}{2}mv^2 = \frac{e^2}{8\pi\epsilon_0 r}$$

$$\therefore E = \frac{e^2}{8\pi\epsilon_0 r} - \frac{e^2}{4\pi\epsilon_0 r} = -\frac{e^2}{8\pi\epsilon_0 r} \quad (\text{v})$$

If the electron can behave as a *wave*, it must be possible to fit a whole number of wavelengths around the orbit. In this case a standing wave pattern is set up and the energy in the wave is confined to the atom. A progressive wave would imply that the electron is moving from the atom and is not in a stationary orbit.

If there are n waves in the orbit and λ is the wavelength,

$$n\lambda = 2\pi r \quad (\text{vi})$$

$$\therefore \lambda = \frac{h}{p} = \frac{h}{mv} \quad (\text{vii})$$

Hence, from (vi) and (vii),

$$\frac{nh}{2\pi} = mvr \quad (\text{viii})$$

Now $mv \times r$ is the moment of momentum or *angular momentum* of the electron about the nucleus. Thus equation (viii) states that the *angular momentum is a multiple of $h/2\pi$* . This quantization of angular momentum, a key point in atomic theory, was first proposed by Bohr in 1913, twelve years before de Broglie proposed the wave-particle relation $\lambda = h/p$.

From (iv), $mv^2 r = e^2/4\pi\epsilon_0$.

Hence, with (viii),
$$v = \frac{2\pi e^2}{4\pi\epsilon_0 nh} = \frac{e^2}{2\epsilon_0 nh}$$

and thus
$$r = \frac{\epsilon_0 n^2 h^2}{\pi m e^2} \quad (\text{ix})$$

In classical physics, charges undergoing acceleration emit radiation

and, therefore, lose energy (see p. 986). On this basis the electron would spiral towards the nucleus and the atom would collapse. Bohr, therefore, suggested (a) that in those orbits where the angular momentum is a multiple of $h/2\pi$ the energy is constant, (b) that the electron, or atom, can pass from one allowed energy level E_1 to another E_2 of smaller value, but not to a value between, and that the difference in energy is released in the form of radiation of energy $h\nu$, where ν is the frequency of the radiation emitted. Thus,

$$E_1 - E_2 = h\nu \quad . \quad . \quad . \quad . \quad (x)$$

From (v) and (ix),

$$\begin{aligned} \therefore E &= -\frac{e^2}{8\pi\epsilon_0 r} = -\frac{e^2}{8\pi\epsilon_0} \times \frac{\pi m e^2}{\epsilon_0 n^2 h^2} \\ &= -\frac{m e^4}{8 \epsilon_0^2 n^2 h^2} \\ \therefore E_1 - E_2 &= \frac{m e^4}{8 \epsilon_0^2 h^2} \left(\frac{1}{n_2^2} - \frac{1}{n_1^2} \right) = h\nu \end{aligned}$$

$$\therefore \nu = \frac{m e^4}{8 \epsilon_0^2 h^3} \left(\frac{1}{n_2^2} - \frac{1}{n_1^2} \right)$$

$$\text{or} \quad \frac{1}{\lambda} = \frac{\nu}{c} = \bar{\nu} = \frac{m e^4}{8 \epsilon_0^2 c h^3} \left(\frac{1}{n_2^2} - \frac{1}{n_1^2} \right) \quad . \quad . \quad . \quad (xi)$$

where λ is the wavelength and $\bar{\nu}$, the *wave number*, is the number of wavelengths per metre.

Spectral Series of Hydrogen

Before Bohr's theory of the hydrogen atom it had been found that the wave numbers of the hydrogen spectrum could be arranged in the form of a series, named after its discoverer. Among the wave numbers were:

$$1. \text{ Lyman (ultra-violet) series—} \bar{\nu} = R \left(\frac{1}{1^2} - \frac{1}{m^2} \right).$$

$$2. \text{ Balmer (visible) series—} \bar{\nu} = R \left(\frac{1}{2^2} - \frac{1}{m^2} \right).$$

$$3. \text{ Paschen (infra-red) series—} \bar{\nu} = R \left(\frac{1}{3^2} - \frac{1}{m^2} \right),$$

where R is a constant known as *Rydberg's constant* and m is an integer. From Bohr's formula in (xi), it follows that all the spectral series can

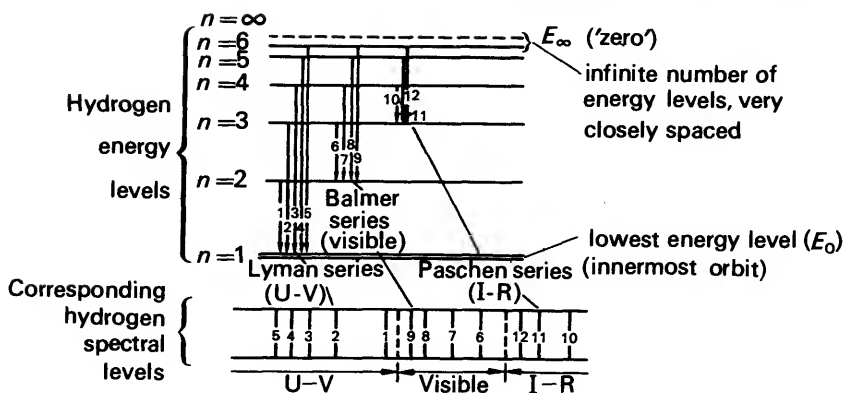


FIG. 43.18. Energy levels and spectra of hydrogen (not to scale).

be obtained simply by putting $n_2 = 1, 2, 3$ respectively and $n_1 = m$ (see Fig. 43.18). Moreover, (a) the agreement between the experimental and theoretical values of the wave numbers is excellent, (b) Rydberg's constant determined experimentally is $1.09678 \times 10^7 \text{ m}^{-1}$ and from $R = me^4/8\epsilon_0^2ch^3$ it is $1.09700 \times 10^7 \text{ m}^{-1}$, (c) the radius of the first circular orbit calculated from $r = \epsilon_0 h^2/\pi me^2$, equation (ix), is $5.29 \times 10^{-11} \text{ m}$. This radius, which corresponds to $n = 1$ and is called the 'first Bohr radius', is the radius of the stable hydrogen atom, since the energy E is a minimum when $n = 1$, from $E = -me^4/8\epsilon_0^2n^2h^2$. The value of r is in good agreement with the atomic radius calculated from the kinetic theory of gases.

Excitation and Ionization Potentials

Bohr's theory of the hydrogen atom was unable to predict the energy levels in complex atoms, which had many electrons. Quantum or wave mechanics, beyond the scope of this book, is used to explain the spectral frequencies of these atoms. The fundamental ideas of Bohr's theory, however, are still retained, for example, the angular momentum of the electron has quantum values and the energy levels of the atom have only allowed discrete values.

Generally, an atom is most stable when it has a minimum energy E_0 , and it is then said to be in its *ground state* (p. 1084). If the atom absorbs energy, and the energy of the atom reaches one of its allowed values E_1 , the atom is said to be in an *excited state*.

The energy required to raise an atom from its ground state to an excited state is called the *excitation energy* of the atom. If the energy is eV , where e is the electronic charge, V is known as the *excitation potential* of the atom.

If an atom is in its ground state with energy E_0 , and absorbs an amount of energy $e\bar{V}$ which just removes an electron completely from the atom, then \bar{V} is said to be the *ionization potential* of the atom. The potential energy of the atom is here denoted by E_∞ , as the ejected electron is so far away from the attractive influence as to be, in effect, at infinity.

E_∞ is taken as the 'zero' energy of the atom, and its other values are thus negative. The ionization potential is given by $E_\infty - E_0 = e\bar{V}$, or by $-E_0 = eV$. Fig. 43.19 shows roughly the energy levels of an atom,

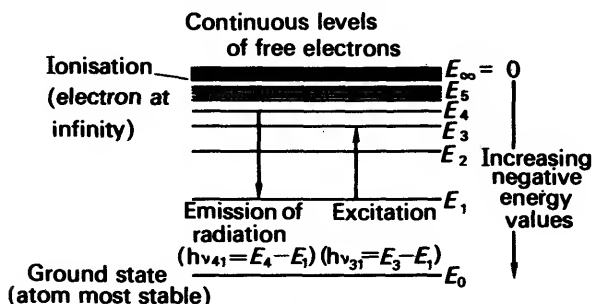


FIG. 43.19. Energy levels in the atom.

namely, its ground state E_0 , its excited states, E_1, E_2, \dots , and its ionization state, E_∞ . It will be noted that the energy levels become more closely spaced at the higher excited states. Beyond the value E_∞ , when the ejected electron is no longer under the attractive influence of the nucleus, the energy of the 'free' electron, $\frac{1}{2}mv^2$, can have one of a continuous range of energies, whereas inside the atom it could only have one of a number of separated allowed values.

Optical Spectra

As an illustration of Bohr's theory of energy levels, the ionization potential, E_∞ , of helium is 24.6 eV (electron-volt). The ground state thus corresponds to an energy level of E_0 of -24.6 eV. Suppose there is an excitation level, E_n , of helium of -21.4 eV. Then if the helium atom is excited to this level and falls directly to the ground state, the frequency ν_n of the radiation emitted is given by

$$h\nu_n = E_n - E_0$$

$$\therefore \nu_n = \frac{[(-21.4) - (-24.6)] \times 1.6 \times 10^{-19}}{6.6 \times 10^{-34}} \text{ Hz,}$$

since h = Planck's constant = 6.6×10^{-34} Js and 1 eV = 1.6×10^{-19} J. Thus the wavelength, λ_n , is given by

$$\lambda_n = \frac{c}{\nu_n} = \frac{3 \times 10^8 \times 6.6 \times 10^{-34}}{3.1 \times 1.6 \times 10^{-19}}$$

$$= 3.9 \times 10^{-7} \text{ m.}$$

This is a wavelength in the violet end of the spectrum. Fig. 43.19 illustrates the emission of radiation as the energy of the atom falls from one level to another.

Emission spectra are classified into *continuous*, *line* and *band* spectra. With few exceptions, incandescent solids and liquids produce a continuous spectrum, one in which all wavelengths are found over a wide range. Line spectra are obtained from *atoms* in gases such as hydrogen

in a discharge tube, and the spectrum of a sodium salt vaporized in a Bunsen flame consists of two lines close together. Gases such as carbon dioxide in a discharge tube also produce a band spectrum, each band consisting of a series of lines very close together at the sharp edge or head of the band and farther apart at the other end or tail. Band spectra are essentially due to *molecules*. The different band heads in a band system are due to small allowed discrete energy changes in the vibrational states of the molecule. The fine lines in a given band are due to still smaller allowed discrete energy changes in the rotational states of the molecule.

Limitations of Bohr's Theory—Quantum Theory

The Bohr theory predicts successfully the energy levels of the hydrogen atom, as we have just seen. There are, however, some details of the spectrum which are not explained. When the radiating atoms are placed in a magnetic field each energy level splits into a number of energy levels near to the main energy level. This causes the spectral lines to split and is known as the *Zeeman effect*. The full quantum theory predicts this behaviour and utilizes four quantum numbers:

1. *Principal quantum number, n* . This specifies the main energy level and corresponds to the number n in the Bohr theory.

2. *Orbital quantum number, l* . This can take values $0, 1, \dots, n-1$, and specifies the angular momentum of the electron. In the Bohr theory, one value of angular momentum corresponds to one value of energy. In quantum theory there are n values of l corresponding to the principal quantum number n . For $n = 1$, the ground state, quantum theory states that the angular momentum is zero, unlike the Bohr theory.

3. *Magnetic quantum number, m* . This can take values $-l, -(l-1) \dots 0, 1, \dots +l$. It specifies the orientation of the electron orbitals in a magnetic field and plays an important part in explaining the Zeeman effect.

4. *Spin quantum number, m_s* . It has already been mentioned that electrons are *spinning* and have a magnetic moment because of this spin.

The two possible values of m_s , $\pm \frac{1}{2}$, specify whether the spin is aligned with, or counter to, an applied magnetic field.

A full use of these quantum numbers gives a more accurate description of the fine details of the hydrogen spectrum.

Pauli Exclusion Principle

Pauli's exclusion principle is needed in describing the electron configuration of multi-electron atoms. It states that *no two electrons in an atom can have the same set of four quantum numbers*. In the lowest energy level, $n = 1$ and hence $l = 0$. This in turn restricts the value of m to zero. m_s , however, may take either of its two values $\pm \frac{1}{2}$ and so there are two electron states with $n = 1$. These two electrons form a *shell*. The two electrons in the shell have energy levels which are very near to one another. This first shell is called the *K shell*.

When $n = 2$, l may take the value 0 or 1 and m the value $-1, 0$, or 1 . The following set of quantum numbers are, therefore, possible:

$$\begin{array}{ll} (2 & 0 & 0 & +\frac{1}{2}), & (2 & 0 & 0 & -\frac{1}{2}). \\ (2 & 1 & -1 & +\frac{1}{2}), & (2 & 1 & -1 & -\frac{1}{2}). \\ (2 & 1 & 0 & +\frac{1}{2}), & (2 & 1 & 0 & -\frac{1}{2}), \\ (2 & 1 & +1 & +\frac{1}{2}), & (2 & 1 & +1 & -\frac{1}{2}). \end{array}$$

There are thus 8 electron states with $n = 2$ and these form the L shell. It is left as an exercise to the reader to show that, for $n = 3$, there are 18 electrons. These form the M shell.

On this basis, hydrogen has 1 electron in the K shell; helium, $Z = 2$, has 2 electrons in the K shell. The maximum number of electrons in the K shell is 2, and hence lithium, $Z = 3$, has 2 electrons in the K shell and 1 electron in the L shell. As the atomic number increases, the electrons fill up the L shell. Fluorine, $Z = 9$, has 2 electrons in the K shell and 7 electrons in the L shell. Neon, $Z = 10$, has 2 electrons in the K shell and 8 electrons in the L shell, the maximum possible. Sodium, $Z = 11$, has 2 electrons in the K , 8 in the L , and 1 in the M shells. Chlorine, $Z = 17$, has 2 electrons in the K , 8 in the L , and 7 in the M shells. The sodium atom has thus 1 'available' electron in its outermost shell, that is, it is an electron donor; on the other hand, the chlorine atom can accommodate 1 more electron in its outermost shell, that is, it is an electron acceptor. When the very stable compound sodium chloride is formed, the outer electron in the sodium atom passes to the chlorine atom. The outermost shells of each atom are now complete, which is a very stable electron arrangement (see p. 848).

The chemical activity, or inactivity, of all elements is explained by their electron shell arrangement. Thus fluorine ($K = 2, L = 7$ electrons), chlorine ($K = 2, L = 8, M = 7$ electrons) and bromine ($K = 2, L = 8, M = 8, N = 7$ electrons) have each a vacancy for 1 electron in their respective outermost shells to make them complete, and have similar chemical properties. Lithium ($K = 2, L = 1$ electron), sodium ($K = 2, L = 8, M = 1$ electron) and potassium ($K = 2, L = 8, M = 8, N = 1$ electron) are all electron donors and have similar chemical activity. In contrast, helium ($K = 2$ electrons), and neon ($K = 2, L = 8$ electrons) have complete shells of electrons and are therefore chemically inactive.

The total number of electrons in all the shells is equal to Z , the atomic number, which is 1 for hydrogen and 92 for uranium.

EXERCISES 43

1. Describe the properties of X-rays and compare them with those of ultra-violet radiation. Outline the evidence for the wave nature of X-rays.

The energy of an X-ray photon is $h\nu$ joules where $h = 6.63 \times 10^{-34}$ J s and ν is the frequency in hertz (cycles per second). X-rays are emitted from a target bombarded by electrons which have been accelerated from rest through 10^5 V. Calculate the minimum possible wavelength of the X-rays assuming that the corresponding energy is equal to the whole of the kinetic energy of one electron. (Charge of an electron = 1.60×10^{-19} C; velocity of electromagnetic waves in *vacuo* = 3.00×10^8 m s $^{-1}$.) (O. & C.)

2. In an experiment on the photoelectric effect using radiation of wavelength 4.00×10^{-7} m the maximum electron energy was observed to be 1.40×10^{-19} joule. With radiation of wavelength 3.00×10^{-7} m the maximum energy was 3.06×10^{-19} joule. Derive a value for Planck's constant.

Mention one other physical phenomenon involving Planck's constant. (Velocity of light = 3.00×10^8 m s $^{-1}$.) (N.)

3. Describe briefly experiments to demonstrate *three* of the following:

(a) That gases absorb strongly some of the characteristic radiations of their emission spectra.

(b) That light waves are transverse.

(c) That X-radiation does not consist of electrically charged particles.

(d) That X-radiation is more strongly absorbed by a sheet of lead than by a sheet of aluminium of the same thickness. (O. & C.)

4. Describe a modern form of X-ray tube and explain its action.

Outline the evidence for believing (a) that X-rays are an electromagnetic radiation, (b) that wavelengths in the X-ray region are of the order of 10^{-3} times those of visible light. (O.)

5. Write down Einstein's equation for photoelectric emission. Explain the meaning of the terms in the equation and discuss its significance.

Describe briefly how Einstein's equation may be verified experimentally.

An effective point source emits monochromatic light of wavelength 4500 Å at a rate of 0.11 watt. How many photons leave the source per second? Light from the source is emitted uniformly in all directions and falls normally on the cathode of area π cm 2 of a photocell at a distance of 50 cm from the source. Calculate the photoelectric current, assuming 10 per cent of the photons incident on the cathode liberate electrons. (Planck's constant = 6.6×10^{-34} J s; charge of electron = 1.6×10^{-19} C.) (O. & C.)

6. Describe the atomic processes in the target of an X-ray tube whereby X-ray line spectra are produced. Determine the ratio of the energy of a photon of X-radiation of wavelength 1 Å to that of a photon of visible radiation of wavelength 5000 Å. Why is the potential difference applied across an X-ray tube very much greater than that applied across a sodium lamp producing visible radiation? (N.)

7. Draw a labelled diagram of some form of X-ray tube and of its electrical connexions when in actual use.

Electrons starting from rest and passed through a potential difference of 1000 volts are found to acquire a velocity of 1.88×10^7 m s $^{-1}$. Calculate the ratio of the charge to the mass of the electron in coulombs per kg. (N.)

8. You are provided with a glass tube containing an electrode at each end, an exhaust pump, and a source of high potential. Under what conditions (a) does

the gas within the tube become a relatively good conductor, (b) is a beam of electrons (cathode rays) produced within the tube?

What modifications must be made in the tube in order that a strong beam of X-rays may be produced? What would happen in each case if the potential applied was increased?

What experiment would you perform to show the effect of a magnetic field on the conducting particles in (a) and (b) and on a beam of X-rays? State the result you would expect. (L.)

9. When light is incident in a metal plate electrons are emitted only when the frequency of the light exceeds a certain value. How has this been explained?

The maximum kinetic energy of the electrons emitted from a metallic surface is 1.6×10^{-19} joule when the frequency of the incident radiation is 7.5×10^{14} Hz. Calculate the minimum frequency of radiation for which electrons will be emitted. Assume that Planck's constant = 6.6×10^{-34} J s. (N.)

10. Describe a modern form of X-ray tube and explain briefly the energy changes that take place while it is operating.

Calculate the energy in electron-volts of a quantum of X-radiation of wavelength 1.5 \AA .

An X-ray tube is operated at 8000 volts. Why is there no radiation of shorter wavelength than that calculated above (while there is a great deal of longer wavelength)? When the voltage is increased considerably above this value why does the spectrum of the radiation include a few (*but only a few*) sharp strong lines, the wavelengths of which depend on the material of the target? (Take $e = 1.6 \times 10^{-19}$ C; $h = 6.5 \times 10^{-34}$ J s; $c = 3 \times 10^8$ m s $^{-1}$.) (O.)

11. How are X-rays produced, and how are their wavelengths determined?

Discuss briefly the origin of the lines in the spectrum produced by an X-ray tube that are characteristic of the target metal.

Give a brief account of Moseley's work and the part it played in establishing the idea of atomic number. (O.)

12. Describe and explain one experiment in which light exhibits a wave-like character and one experiment which illustrates the existence of photons.

Light of frequency 5.0×10^{14} Hz liberates electrons with energy 2.31×10^{-19} joule from a certain metallic surface. What is the wavelength of ultra-violet light which liberates electrons of energy 8.93×10^{-19} joule from the same surface? (Take the velocity of light to be 3.0×10^8 m s $^{-1}$, and Planck's constant (h) to be 6.62×10^{-34} J s.) (L.)

13. What are the chief characteristics of a line spectrum? Explain how line spectra are used in analysis for the identification of elements.

Fig. 43.20, which represents the lowest energy levels of the electron in the

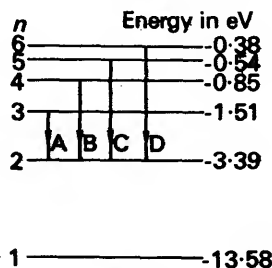


FIG. 43.20.

hydrogen atom, specifies the value of the principal quantum number n associated with each state and the corresponding value of the energy of the level, measured in electron volts. Work out the wavelengths of the lines associated with the transitions A, B, C, D marked in the figure. Show that the other transitions that can occur give rise to lines that are either in the ultra-violet or the infra-red regions of the spectrum. (Take 1 eV to be $1.6 \times 10^{-19} \text{ J}$; Planck's constant h to be $6.5 \times 10^{-34} \text{ J s}$; and c , the velocity of light *in vacuo*, to be $3 \times 10^8 \text{ m s}^{-1}$.) (O.)

14. Einstein's equation for the photoelectric emission of electrons from a metal surface can be written $h\nu = \frac{1}{2}mv^2 + \phi$, where ϕ is the work function of the metal, and consistent energy units are used for each term in the equation. Explain briefly the physical process that this equation represents. Outline an experiment by which you could determine the values of h (or of h/e) and ϕ .

For caesium the value of ϕ is 1.35 electron-volts. (a) What is the longest wavelength that can cause photo-electric emission from a caesium surface? (b) What is the minimum velocity with which photoelectrons will be emitted from a caesium surface illuminated with light of wavelength 4000 \AA ? (c) What potential difference will just prevent a current passing through a caesium photocell illuminated with light of 4000 \AA wavelength? (O.)

Summary of C.G.S. and SI units

QUANTITY AND SYMBOL	C.G.S UNIT	SI UNIT	RELATIONSHIP
mass (m)	gramme (g)	kilogramme (kg)	1000 g = 1 kg
length (l)	centimetre (cm)	metre (m)	100 cm = 1 m
time (t)	second (s)	second (s)	

MECHANICS, FLUIDS

Linear motion			
velocity (v)	cm s^{-1}	m s^{-1}	$1 \text{ cm s}^{-1} = 10^{-2} \text{ m s}^{-1}$
acceleration (a)	cm s^{-2}	m s^{-2}	$1 \text{ cm s}^{-2} = 10^{-2} \text{ m s}^{-2}$
force (F)	dyne (dyn)	newton (N)	$1 \text{ dyn} = 10^{-5} \text{ N}$
momentum (p)	g cm s^{-1}	kg m s^{-1}	$1 \text{ g cm s}^{-1} = 10^{-5} \text{ kg m s}^{-1}$
work, energy (W)	erg	joule (J)	$1 \text{ erg} = 10^{-7} \text{ J}$
power (P)	erg s^{-1}	watt (W)	$1 \text{ erg s}^{-1} = 10^{-7} \text{ W}$
Rotational motion			
angular velocity (ω)	rad s^{-1}	rad s^{-1}	
angular acceleration ($d\omega/dt$)	rad s^{-2}	rad s^{-2}	
moment of inertia (I)	g cm^2	kg m^2	$1 \text{ g cm}^2 = 10^{-7} \text{ kg m}^2$
couple or torque (T)	dyn cm	Nm	$1 \text{ dyn cm} = 10^{-7} \text{ Nm}$
angular momentum (L)	$\text{g cm}^2 \text{ s}^{-1}$	$\text{kg m}^2 \text{ s}^{-1}$	$1 \text{ g cm}^2 \text{ s}^{-1} = 10^{-7} \text{ kg m}^2 \text{ s}^{-1}$
Fluids			
volume (V)	cm^3	m^3	$1 \text{ cm}^3 = 10^{-6} \text{ m}^3$
density (ρ)	g cm^{-3}	kg m^{-3}	$1 \text{ g cm}^{-3} = 10^3 \text{ kg m}^{-3}$
pressure (p)	dyn cm^{-2}	N m^{-2}	$1 \text{ dyn cm}^{-2} = 10^{-1} \text{ N m}^{-2}$

length 1 micron (μm) = 10^{-6} m ; 1 nanometre (nm) = 10^{-9} m ; 1 Angstrom unit, \AA = 10^{-10} m = 10^{-1} nm
time 1 millisecond (ms) = 10^{-3} s ; 1 microsecond (μs) = 10^{-6} s ; 1 nanosecond (ns) = 10^{-9} s
volume 1 c.c. = 1 cm^3 = 10^{-6} m^3 ; 1 litre = 10^{-3} m^3
density water (ρ_w) = 1000 kg m^{-3} ; mercury (ρ_{Hg}) = $13\,600 \text{ kg m}^{-3}$; air at s.t.p. (ρ_{air}) = 1.29 kg m^{-3}

gravitational $g = 981 \text{ cm s}^{-2} = 9.81 \text{ m s}^{-2} = 10 \text{ m s}^{-2}$ (approx.)
force $G = 6.7 \times 10^{-8} \text{ dyn cm}^2 \text{ g}^{-2} = 6.7 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$ (approx.)
power $1 \text{ kgf} = 9.8 \text{ N}$ (approx.) $= 10 \text{ N}$ (approx.)
pressure $1 \text{ h.p.} = 746 \text{ W} = \frac{3}{4} \text{ kW}$ (approx.)
 $1 \text{ bar} = 10^6 \text{ dyn cm}^{-2} = 10^5 \text{ N m}^{-2} = 750 \text{ torr (mmHg)}$ (approx.)
 standard atmospheric pressure $= 760 \text{ torr} = 1.01325 \times 10^5 \text{ N m}^{-2}$ (exact)

PROPERTIES OF MATTER

QUANTITY AND SYMBOL	C.G.S. UNIT	SI UNIT	RELATIONSHIP
surface tension coefficient (γ)	dyn cm^{-1}	N m^{-1}	$1 \text{ dyn cm}^{-1} = 10^{-3} \text{ N m}^{-1}$
modulus of elasticity (E, K or G)	dyn cm^{-2}	N m^{-2}	$1 \text{ dyn cm}^{-2} = 10^{-1} \text{ N m}^{-2}$
viscosity coefficient (η)	poise	N s m^{-2}	$1 \text{ dyn s cm}^{-2} = 10^{-1} \text{ N s m}^{-2}$

γ (water-air at 15°C) $= 74 \text{ dyn cm}^{-1} = 7.4 \times 10^{-2} \text{ N m}^{-1}$
 E (steel) $= 2.1 \times 10^{12} \text{ dyn cm}^{-2} = 2.1 \times 10^{11} \text{ N m}^{-2}$
 η (air at 15°C) $= 1.8 \times 10^{-4} \text{ poise} = 1.8 \times 10^{-5} \text{ N s m}^{-2}$

HEAT

quantity of heat (Q)	calorie (cal)	joule (J) kilojoule (kJ)	$1 \text{ cal} = 4.2 \text{ J}$ ($4.1868 \text{ J} = 1 \text{ cal}_\text{H}$) $1 \text{ kcal} = 4.2 \text{ kJ}$ $T(\text{K}) = 273.15 + t(^{\circ}\text{C})$ $1 \text{ deg C} = 1 \text{ K}$ $1 \text{ cal deg C}^{-1} = 4.2 \text{ J K}^{-1}$ $1 \text{ cal g}^{-1} \text{ deg C}^{-1} = 4.2 \text{ kJ kg}^{-1} \text{ K}^{-1}$ $1 \text{ cal g}^{-1} = 4.2 \text{ kJ kg}^{-1}$ $\text{deg C}^{-1} = \text{K}^{-1}$ $1 \text{ cal g}^{-1} \text{ }^{\circ}\text{K}^{-1} = 4.2 \text{ kJ kg}^{-1} \text{ K}^{-1}$ $1 \text{ cal g-mol}^{-1} \text{ }^{\circ}\text{K}^{-1} = 4.2 \text{ J mol}^{-1} \text{ K}^{-1}$ $1 \text{ cal s}^{-1} \text{ cm}^{-1} \text{ }^{\circ}\text{K}^{-1} = 420 \text{ W m}^{-1} \text{ K}^{-1}$ $1 \text{ cal s}^{-1} = 4.2 \text{ W}$
temperature (t)	kilocal $^{\circ}\text{C}$	K	
temperature change (t)	$\text{deg C } (^{\circ}\text{C})$	K	
heat capacity (C)	cal deg C^{-1}	J K^{-1}	
specific heat capacity (c)	$\text{cal g}^{-1} \text{ deg C}^{-1}$	$\text{J kg}^{-1} \text{ K}^{-1}$	
specific latent heat (l)	cal g^{-1}	J kg^{-1}	
linear, cubic expansivity (α, γ)	deg C^{-1}	K^{-1}	
gas constant per unit mass	$\text{cal g}^{-1} \text{ }^{\circ}\text{K}^{-1}$	$\text{J kg}^{-1} \text{ K}^{-1}$	
molar gas constant	$\text{cal g-mol}^{-1} \text{ }^{\circ}\text{K}^{-1}$	$\text{J mol}^{-1} \text{ K}^{-1}$	
thermal conductivity (k)	$\text{cal s}^{-1} \text{ cm}^{-1} \text{ }^{\circ}\text{C}^{-1}$	$\text{W m}^{-1} \text{ K}^{-1}$	
radiation (heat per second)	cal s^{-1}	W	

	WATER (approximate values)	COPPER (approximate values)
	C.G.S.	C.G.S.
specific heat capacity	$1 \text{ cal g}^{-1} \text{ deg C}^{-1}$	$0.095 \text{ cal g}^{-1} \text{ deg C}^{-1}$
specific latent heat of fusion	80 cal g^{-1}	49 cal g^{-1}
specific latent heat of vaporisation	540 cal g^{-1}	206 kJ kg^{-1}
thermal conductivity (20°C)	$1.4 \times 10^{-3} \text{ cal s}^{-1} \text{ cm}^{-1} \text{ }^\circ\text{C}^{-1}$	1250 kJ kg^{-1}
expansivity (20°C)	$(\gamma) 2 \times 10^{-4} \text{ deg C}^{-1}$	$0.91 \text{ cal s}^{-1} \text{ cm}^{-1} \text{ }^\circ\text{C}^{-1}$
		$(\alpha) 17 \times 10^{-6} \text{ deg C}^{-1}$
	SI	SI
gas constant per unit mass, hydrogen	$1 \text{ cal g}^{-1} \text{ K}^{-1}$	$4.15 \text{ kJ kg}^{-1} \text{ K}^{-1}$
molar gas constant	$2 \text{ cal g-mol}^{-1} \text{ deg C}^{-1}$	$8.3 \text{ J mol}^{-1} \text{ K}^{-1} = 8.3 \text{ kJ kmol}^{-1} \text{ K}^{-1}$
Avogadro constant (N_A or L)	$6.02 \times 10^{23} \text{ g-mol}^{-1}$	$6.02 \times 10^{23} \text{ mol}^{-1} = 6.02 \times 10^{26} \text{ kmol}^{-1}$
Stefan constant (σ)	$5.7 \times 10^{-5} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ K}^{-4}$	$5.7 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$

ELECTRICITY

QUANTITY AND SYMBOL	C.G.S. UNIT	SI UNIT	RELATIONSHIP
electric charge (Q)	e.s.u./e.m.u.	coulomb (C)	1 e.s.u. = $1/(3 \times 10^9) \text{ C}$
permittivity (ϵ)	e.s.u.	F m^{-1}	1 e.s.u. = $8.85 \times 10^{-12} \text{ F m}^{-1}$
electric intensity (E)	e.s.u.	V m^{-1} or N C^{-1}	1 e.s.u. = $3 \times 10^4 \text{ V m}^{-1}$
			1 e.s.u. = 300 V
electric potential, p.d. (V)	e.s.u.	V	1 e.m.u. = 10^{-8} V
capacitance (C)	e.s.u.	F	1 e.s.u. = $1/(9 \times 10^{11}) \text{ F}$
surface density (σ)	e.s.u.	C m^{-2}	1 e.s.u. = $1/(3 \times 10^5) \text{ C m}^{-2}$
current (I)	e.m.u.	A	1 e.m.u. = 10 A
resistance (R)	e.m.u.	Ω	1 e.m.u. = $10^{-9} \Omega$
resistivity (ρ)	$\Omega \text{ cm}$	$\Omega \text{ m}$	1 $\Omega \text{ cm} = 10^{-2} \Omega \text{ m}$
electrochem. equivalent (z)	g C^{-1}	kg C^{-1}	1 $\text{g C}^{-1} = 10^{-3} \text{ kg C}^{-1}$
magnetic flux density (B)	gauss	tesla, T ($= \text{Wb m}^{-2}$)	1 gauss = 10^{-4} T
or magnetic induction			
magnetic flux (Φ)	maxwell	Wb	1 maxwell = 10^{-8} Wb
permeability (μ)	e.m.u.	H m^{-1}	1 e.m.u. = $4\pi \times 10^{-7} \text{ H m}^{-1}$
self inductance (L)	e.m.u.	H	1 e.m.u. = 10^{-9} H
magnetising field strength (H)	oersted (Oe)	A m^{-1}	1 Oe = $(10^3/4\pi) \text{ A m}^{-1}$
magnetic moment (m)	e.m.u.	A m^2	1 e.m.u. = 10^{-3} A m^2

(Approximate values)
permittivity of vacuum, $\epsilon_0 = 8.85 \times 10^{-12} \text{ F m}^{-1}$
permeability of vacuum, $\mu_0 = 4\pi \times 10^{-7} \text{ H m}^{-1}$ (exact)
resistivity of copper $= 1.7 \times 10^{-6} \Omega \text{ cm} = 1.7 \times 10^{-8} \Omega \text{ m}$
e.c.e of hydrogen $= 1.1 \times 10^{-5} \text{ g C}^{-1} = 1.1 \times 10^{-8} \text{ kg C}^{-1}$
earth's horizontal component field $= 0.2 \text{ gauss} = 2 \times 10^{-5} \text{ T}$
moving-coil meter field $= 5000 \text{ gauss} = 0.5 \text{ T}$

OPTICS

QUANTITY AND SYMBOL	C.G.S. UNIT	SI UNIT	RELATIONSHIP
luminous flux (Φ)	lumen (lm)	lumen (lm)	
luminous intensity (I)	c.p.	candela (cd)	1 c.p. = 1 cd (approx.)
illumination (E)	cm-candle	lux	1 cm-candle = 10^4 lux
luminance (L)	lm cm ⁻²	lm m ⁻²	1 lm cm ⁻² = $10^{-4} \text{ lm m}^{-2}$

SOUND

frequency (f)	Hz	Hz	
intensity (I)	W cm ⁻²	W m ⁻²	1 W cm ⁻² = 10^{-4} W m^{-2}

speed of light in vacuo $= 3 \times 10^{10} \text{ cm s}^{-1} = 3 \times 10^8 \text{ m s}^{-1}$
speed of sound in air at 0°C $= 33200 \text{ cm s}^{-1} = 332 \text{ m s}^{-1}$
threshold of hearing $= 10^{-16} \text{ W cm}^{-2} = 10^{-12} \text{ W m}^{-2}$
wavelength of sodium (yellow) light $= 5.89 \times 10^{-5} \text{ cm} = 5.89 \times 10^{-7} \text{ m} = 589 \text{ nm}$

SOME ATOMIC CONSTANTS (*approximate values*)

	C.G.S.	SI
electron charge (e)	4.8×10^{-10} e.s.u.; 1.6×10^{-20} e.m.u.	1.6×10^{-19} C
electron rest mass (m_e)	9.1×10^{-28} g	9.1×10^{-31} kg
electron specific charge (e/m_e)	5.3×10^{17} e.s.u. g $^{-1}$; 1.76×10^7 e.m.u. g $^{-1}$	1.76×10^{11} C kg $^{-1}$
proton rest mass (m_p)	1.7×10^{-24} g	1.7×10^{-27} kg
proton specific charge (e/m_p)	9600 e.m.u. g $^{-1}$	9.6×10^7 C kg $^{-1}$
speed of e-m waves (c)	3.0×10^{10} cm s $^{-1}$	3.0×10^8 m s $^{-1}$
Planck constant (h)	6.6×10^{-27} erg s	6.6×10^{-34} J s
electron-volt (eV)	1.6×10^{-12} erg	1.6×10^{-19} J
atomic mass unit (a.m.u.)	1.66×10^{-24} g	1.66×10^{-27} kg

References

- Symbols, Signs and Abbreviations*—The Royal Society.
Physico-Chemical Quantities and Units—Prof. M. L. McGlashan, Royal Institute of Chemistry.
The Use of SI Units—PD5686—British Standards Institution.
Changing to the Metric System—National Physical Laboratory—HMSO.
Signs, Symbols and Abbreviations—Association for Science Education.

Answers to Exercises

MECHANICS

EXERCISES 1 (p. 32)

1. LT^{-1} . 2. MLT^{-2} . 3. (i) scalar, (ii) vector, (iii) scalar, (iv) vector. 4. mass \times velocity. 5. momentum, energy. 6. total linear momentum, external. 7. 1 joule. 8. 10 N (approx.). 9. vector. 10. rate. 11. C. 12. B. 13. D. 14. A. 15. (i) 5 s, (ii) 62.5 m, (iii) 18 m s $^{-1}$. 16. (i) 4 s, (ii) 20 m, (iii) 10, 10 J. 17. 19.8 km h $^{-1}$, N. 30.5° W. 18. (i) 100, (ii) 500 J. 19. (i) 5.2 m s $^{-1}$, loss = 58 J, (ii) 1.2 s $^{-1}$, loss = 314 J. 20. 26.5 km h $^{-1}$, S. 41° W., 7.6 km. 21. (i) 1 $\frac{3}{4}$ m s $^{-1}$, (ii) 1 $\frac{3}{4}$ m s $^{-2}$, (iii) 86 J. 22. 10.5° from vertical. 23. 2E/103. 24. (a) 10/3 N, (b) 5/9 W, (c) 5/18 W. 25. 167 kgf; 83330 J. 26. 21675 m, 3,125 m, 25 s; 465 m s $^{-1}$. 27. 1/3. 28. (a) ML^2T^{-2} , (b) a: ML^5T^{-2} , b: L^3 . 29. 14.4 minutes, 8 km, 37° S. of E. 30. $v/2$ at 60° to initial velocity of first sphere. 31. 2 m s $^{-1}$. 32. 800 kgf, 22 kW. 33. 70% 0.023. 34. 22.5×10^4 N m $^{-2}$.

EXERCISES 2 (p. 69)

1. centripetal. 2. middle. 3. $g = GM/r^2$. 4. end. 5. 1/distance 2 . 6. 24 h. 7. B. 8. D. 9. C. 10. D. 11. (i) 2 rad s $^{-1}$, (ii) 9.6 kgf. 12. (i) 11.8 kgf, (ii) 32°. 13. 42°, 1555 kgf. 14. 22.4, 6.4 kgf. 15. (a) mg , (b) $\sqrt{2}gl$, (c) $2g$ up, (d) 3 mg . 16. Break when stone vertically below point of suspension; 7.7 rad s $^{-1}$; 122 cm from point below point of suspension. 17. (i) 1/20 s, (ii) 0, 3200 π^2 , (iii) $40\sqrt{3}\pi$, 0. 18. 101 cm, (i) 0, $2\pi^2$ cm s $^{-2}$, (ii) 2π cm s $^{-1}$, 0, (iii) $\sqrt{3}\pi$ cm s $^{-1}$, π^2 cm s $^{-2}$. 19. 0.32 s. 20. 1.6 s. 21. 10 m s $^{-2}$, 4.5 m. 22. (a) 2.5 m s $^{-1}$, (b) 790 m s $^{-2}$. 23. 6.3 s. 26. 1.6 Hz. 27. (a) $4\pi^2$, (b) $2\sqrt{3}\pi^2$, (c) $4\pi^3$; $16\pi^4mr$. 30. $\frac{1}{2}m\omega^2(a^2 - x^2)$. 32. 5×10^{-2} m. 33. 3×10^{-7} N. 34. 6.0×10^{24} kg. 35. 9.9 m s $^{-2}$. 38. 1.4×10^{-5} rad s $^{-1}$. 39. 9.77 m s $^{-2}$. 40. 5500 kg m $^{-3}$. 42. 24 hours.

EXERCISES 3 (p. 93)

1. $\frac{1}{2}I\omega^2$. 2. $I\omega$. 3. couple. 4. $2\pi\sqrt{1/mgh}$. 5. D. 6. B. 7. A. 8. E. 9. (i) 2×10^{-4} , (ii) 8×10^{-4} kg m 2 . 10. (i) 8, (ii) 24 kg m 2 . 11. (i) 4×10^{-3} , (ii) 8×10^{-3} , (iii) 2×10^{-3} kg m 2 . 12. 15 joule. 13. 3.6 m s $^{-2}$, 6.0 m s $^{-1}$. 14. (i) 1.9, (ii) 2.2 rad s $^{-1}$. 15. 4.0 s. 16. (a) 20 rad s $^{-2}$, (b) 0.32 N m. 17. 8.1×10^{-4} kg m 2 . 18. 4.02×10^{-3} kg m 2 . 19. 25Mh $^2/7$. 20. 3.5 m s $^{-1}$. 21. 6.2×10^{-4} kg m 2 . 23. 1:12:5.

EXERCISES 4 (p. 121)

1. newton metre. 2. rises. 3. meet. 4. $F \cos \theta$. 5. centre of gravity. 6. weight of object. 7. velocity. 8. B. 9. D. 10. C. 11. 22.6 cm. 12. 69,139 kgf. 13. 50 kgf m; 16.7 kgf. 14. (i) 0.65, (ii) 0.60 m. 15. $4\frac{1}{6}$, $4\frac{1}{6}$ kgf. 16. 30°. 18. 171 kgf. 19. 0.04 cm; 0.1°. 20. $\pm \cos^{-1}[(M+m)r \sin \phi / MI]$. 22. 18, 15.9 kgf. 23. stable if $r > t/2$. 25. (i) 3:2, (ii) 40:9. 26. 162 gf. 27. 5.1 cm. 28. 2:3. 29. $W\rho/\sigma$, (a) 0.95, (b) 1.19. 30. (a) 75.0 cm, (b) 1003.3 cm. 33. 1.25. 34. 5.7 cm 3 .

PROPERTIES OF MATTER

EXERCISES 5 (p. 148)

1. N m^{-1} . 2. MT^{-2} . 3. minimum. 4. $4\gamma/r$. 5. $2\gamma/r$. 6. obtuse. 7. per unit length on one side of a line in the surface. 8. free surface energy. 9. C. 10. B. 11. D. 12. A. 13. 1.4×10^{-2} , 8.68×10^{-3} N. 14. MT^{-2} , (i) 6.6, (ii) 5.5 cm. 15. 2.7 cm. 16. (i) $1.00056 \times 10^5 \text{ N m}^{-2}$, (ii) 0.14 cm. 19. 0.75 cm. 20. 2.6 cm. 21. $8\pi T(b^2 - a^2)$. 22. 7.35 cm, angle of contact now 47° . 25. $\pi r^2 h \rho - 2\pi r \gamma \cos \alpha$. 26. 5.6 cm, angle of contact 44° . 27. 0.18 cm. 28. (a) 3.2 cm, (b) $7.1 \times 10^{-2} \text{ N m}^{-1}$. 30. 0.8. 31. $7.4 \times 10^{-3} \text{ N}$.

EXERCISES 6 (p. 167)

1. tensile. 2. N m^{-2} . 3. elastic limit. 4. $\frac{1}{2}$ load. 5. pressure. 6. shear. 7. C. 8. B. 9. A. 10. D. 11. $6.2 \times 10^6 \text{ N m}^{-2}$, 6×10^{-5} , $1.0 \times 10^{11} \text{ N m}^{-2}$. 12. 6.8 kg. 13. 0.08 mm. 15. 7.13 kgf. 16. 1.2 N. 17. 20 m s^{-1} . 18. 117 gf. 19. 2%. 20. $1/120 \text{ J}$. 21. $1.1 \times 10^4 \text{ J}$. 22. (a) $2.2 \times 10^{-6} \text{ m}$, (b) $5.46 \times 10^{-5} \text{ J}$. 23. 0.5. 24. 144°C , 3360 kg. 25. (a) $2 \times 10^{11} \text{ N m}^{-2}$, (b) $4.7 \times 10^{-2} \text{ J}$. 26. $1.1 \times 10^4 \text{ N}$. 27. $7.8 \times 10^{10} \text{ N m}^{-2}$. 28. 40, 74 N. 29. 28.6 kgf. 30. $12 \times 10^7 \text{ N m}^{-2}$, 3 kgf.

EXERCISES 7 (p. 183)

1. F/R . 2. independent. 3. $\pi p a^4/8\eta l$. 4. $\text{ML}^{-1}\text{T}^{-1}$. 5. $6\pi\eta a v$. 6. volume. 7. C. 8. B. 9. D. 10. A. 11. 4.97 cm. 12. (a) 17° , (b) 36 gf. 13. 0.6. 14. 160 J. 15. (a) 36, (b) 60 rev min^{-1} . 17. 0.5, 13.5 cm. 18. $x = 2$, $y = 1$, $z = 2$; $k = 2.3$. 20. 0.025 m s^{-1} . 23. 17 s.

HEAT

EXERCISES 9 (p. 216)

1. $2.2 \text{ kJ kg}^{-1} \text{ K}^{-1}$ ($2.2 \text{ J g}^{-1} \text{ K}^{-1}$). 2. 351 kJ kg^{-1} (351 J g^{-1}). 3. 378 kJ kg^{-1} (378 J g^{-1}). 4. 904 m. 5. 8.8 g. 6. 74 km h^{-1} . 7. $0.76 \text{ kJ kg}^{-1} \text{ K}^{-1}$ ($0.76 \text{ J g}^{-1} \text{ K}^{-1}$). 30.8 min. 8. 0.68 m. 9. 417 kJ kg^{-1} (417 J g^{-1}), 6.7 W. 11. 2230 kJ kg^{-1} (2230 J g^{-1}). 13. 59.3 kg. 14. 0.010 s. 15. 50 g; 328 kJ kg^{-1} (328 J g^{-1}). 16. 42 J K^{-1} . 17. $2.22 \text{ kJ kg}^{-1} \text{ K}^{-1}$ ($2.22 \text{ J g}^{-1} \text{ K}^{-1}$). 18. 93.5 g. 19. -186°C . 20. $4.2 \text{ kJ kg}^{-1} \text{ K}^{-1}$ ($4.2 \text{ J g}^{-1} \text{ K}^{-1}$).

EXERCISES 10 (p. 264)

1. 0.014 m^3 . 2. 3.75 at. 3. 878 mmHg. 4. (i) 2.08 kJ kg^{-1} , (ii) 842 mmHg. 5. 100°C . 8. 146.5 mmHg. 9. 164 J. 11. $0.53 \text{ kJ kg}^{-1} \text{ K}^{-1}$ ($0.53 \text{ J g}^{-1} \text{ K}^{-1}$); $1/3$. 13. (a) 586 K, (b) 101.2 J , (c) 355 J. 14. $0.725 \text{ kJ kg}^{-1} \text{ K}^{-1}$ ($0.725 \text{ J g}^{-1} \text{ K}^{-1}$). 15. 222 K, 38.4 cmHg. 16. 1.67. 17. 56.8 cmHg, 227 K. 18. 597 m s^{-1} . 21. 0.21 mmHg (27.5 N m^{-2}). 22. 1305 m s^{-1} , 0°C . 26. (b) 0.31 J. 27. 164 J, 30 J.

EXERCISES 11 (p. 291)

1. 79°C . 2. 434°C . 3. 762.4 mmHg. 4. $8.5 \times 10^{-6} \text{ K}^{-1}$. 5. $3.8 \times 10^{-4} \text{ cm}^2$. 6. 964 g. 7. 62.4 g. 8. $12 \times 10^{-6} \text{ K}^{-1}$. 9. 4 s. 10. 300 N. 12. 20.5 cm. 13. 270°C . 14. 0.436, 0.444 cm^3 .

EXERCISES 12 (p. 327)

3. 91.7 torr. 5. 26°C . 7. 4.45 cmHg. 10. 707 mm. 11. air: $1.18 \times 10^3 \text{ kg}$, vapour: 8.96 kg. 13. 78%.

EXERCISES 13 (p. 361)

1. $42 \text{ W m}^{-1} \text{ K}^{-1}$, $0.12 \text{ W m}^{-1} \text{ K}^{-1}$. 2. 89°C = junction temp. 3. (a) 90°C , (b) 2.86 W . 4. 354 J . 5. 1700 g . 6. $41:1$. 7. $2.9 \times 10^{-2} \text{ W}$, $0.36 \text{ kJ kg}^{-1} \text{ K}^{-1}$. 8. 2.8 W . 9. 22 W , 10% . 12. $5.7 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$. 13. 2140 K . 14. Newton: (a) 16 , (b) $136 \text{ degC min}^{-1}$; Stefan: (a) 19 , (b) $1374 \text{ degC min}^{-1}$. 15. $5.6 \times 10^7 \text{ J}$. 17. 5450°C . 19. 40 J m^{-2} , 2 degC min^{-1} . 20. 87600 J . 21. 5490°C .

EXERCISES 14 (p. 381)

1. 68°C , -272°C . 3. 385°C . 4. (a) -274°C , (b) 99.45 cm . 8. 50.4°C . 10. 815 mmHg . 14. 0.89°C . 15. 309°C . 16. 77.5 cmHg .

OPTICS

EXERCISES 16 (p. 400)

4. $4a$; $2na$.

EXERCISES 17 (p. 417)

1. (i) 15 cm , 1.5 , (ii) 12 cm , 3. 3. 6 cm , 0.4 . 6. $4/21 \text{ m}$. 7. Object distance = 10 cm , $r = 40 \text{ cm}$; concave. 9. $2R$. 10. 2 radians , or 114° . 11. Inverted. 12. 4.5 cm behind mirror; 0.25 mm ; $5/38$. 13. (a) 240 cm , (b) 1.3 cm .

EXERCISES 18 (p. 437)

1. 35.3° . 2. 41.8° . 3. (i) 26.3° , (ii) 56.4° . 4. (b) 3 cm from bottom. 6. (i) 41.8 , (ii) 48.8° , (iii) 62.5° . 9. 1.47 . 11. (b) 12 cm above mirror. 15. 1.60 . 17. 1.41 . 20. Nearer by $(n-1)d/n$.

EXERCISES 19 (p. 451)

1. 42° . 2. (i) 1.52 , (ii) 52.2° . 3. 60° , $55^\circ 30'$, 1.648 . 4. $4^\circ 48'$. 5. angle i on second face = 60.7° , $c = 41.8^\circ$. 7. $43^\circ 35'$. 8. $37^\circ 45'$; $10^\circ 8'$; 180° . 10. 55° , 1.53 . 12. 27.9° .

EXERCISES 20 (p. 468)

1. Crown: (i) 3.07° , (ii) 3.14° , (iii) 3.11° ; flint: (i) 2.58° , (ii) 2.66° , (iii) 2.62° . 2. 0.023 , 0.031 . 3. 3.92° , 0.021° . 4. 0.54 mm ; 0.54 mm . 5. 1.75 . 8. 0.144° . 11. (a) $49^\circ 12'$, (b) $50^\circ 38'$, (c) $1^\circ 26'$. 12. 6.67° , 0.83° . 13. 1170 km s^{-1} .

EXERCISES 21 (p. 504)

1. (i) 40 cm virtual, (ii) 80 cm real. 2. 7.2 , 18 cm from nearest point on sphere. 3. (a) 1.51 , (b) 7.5 cm . 4. $v = 6r$, where r is radius. 5. (i) 12 cm , $m = 1$, (ii) 12 cm , $m = 3$. 6. $6\frac{2}{3} \text{ cm}$, $\frac{3}{5}$. 7. $13\frac{1}{3}$, 40 cm . 8. (i) $5\frac{1}{2} \text{ cm}$, (ii) $22\frac{1}{2} \text{ cm}$. 9. (i) $9\frac{3}{5} \text{ cm}$, (ii) $16\frac{2}{3} \text{ cm}$. 10. 10 cm . 11. 1.4 . 12. 12.0 , 18.7 cm . 13. 80 cm . 14. (a) 40 cm , (b) 160 cm . 15. (a) $11\frac{1}{5}$, (b) $5\frac{5}{9} \text{ cm}$; $3\frac{2}{7}$. 16. $27\frac{1}{4} \text{ cm}$ from lens. 18. 1.44 . 19. radii = 9.9 , 24.8 cm ; $n = 1.51$. 20. 4 cm . 21. $12\frac{5}{6}$, $37\frac{5}{6} \text{ cm}$ above water surface. 22. 1.4 . 23. (a) 20.5 cm , (b) 12.85 cm , (c) 1.63 . 24. (a) 120 cm from converging lens, (b) 92.2 cm , (c) 2.2 . 25. (a) Beside object O, (b) 72.5 cm from O.

EXERCISES 22 (p. 522)

1. Diverging, $f = 200 \text{ cm}$; $22\frac{2}{3} \text{ cm}$. 2. Converging, $f = 28\frac{4}{7} \text{ cm}$; $35\frac{5}{7}$ to 25 cm . 3. near pt.: 50 cm , far pt.: 200 cm , $r = 16\frac{2}{3}$, $+100 \text{ cm}$. 5. 220 cm . 6. Diverging,

$f = 20$ cm. Infinity to 20 cm from eye. 8. Diverging, $f = 30$ cm; 30 cm. 9. 15.3 cm; 39.7 cm, diverging. 10. 40 cm; converging. 12. 64 cm; $106\frac{2}{3}$ cm. 14. 0.98, 1.02 cm; 3.8. 15. 10 cm.

EXERCISES 23 (p. 548)

1. 8:5:2. 2. (a) infinity, (b) least distance of distinct vision. 3. (a) 4, (b) 4.8. 4. long sight, $f = +38\frac{8}{9}$ cm. 5. 30 cm from scale; $f = 6$ cm. 8. $f_e = 4$ cm, dia. = 0.5 cm; $r = 4.8$ cm; 0.013 cm. 9. 3.2, 8.8 cm. 10. 4.0, 21.0 cm. $M = 6.0$. 11. (a) 250, (b) 0.12 cm. 12. 89° or 91° . 13. 8.7 cm, 46.7. 14. 0.55 cm; 2:1. 15. 2; 22.5 cm. 16. (a) 22.4, (b) 4.9 cm diameter. 17. 6.25×10^{-3} rad.

EXERCISES 24 (p. 573)

7. 22.5 m; 0.02 m. 8. 25; 6×10^4 . 9. 25 rev s^{-1} , 3×10^4 cycles. 12. 187.5 lux (m-candle), 67.5 cd. 13. (i) $3\frac{1}{8}$, (ii) 1.6 lux. 14. 125 cd. 15. 2, 10 lux. 16. 0.5. 17. 1, 0.83 lux; 1.41, 1.2 lux. 18. 2.63 metres. 19. 64 lumen m^{-2} . 20. 63.6° . 22. 90.3% .

WAVES AND SOUND

EXERCISES 25 (p. 604)

2. $L = \text{sound}$, $T = \text{remainder}$. 3. (i) 133 cm, (ii) 400 Hz. 4. (i) 100 Hz, (ii) 1.7 m, (iii) 170 m s^{-1} , (iv) π , (v) $0.2 \sin(400\pi t + 20\pi x/17)$. 6. 380, 391 Hz. 7. 625.9 Hz. 10. (i) $5\pi/3$, (ii) $y = 0.03 \sin 2\pi(250t - 25x/3)$, (iii) 6 cm, (iv) $0.01 \sin 50\pi x/3 \cdot \cos 500\pi t$ (x in m).

EXERCISES 26 (p. 635)

1. (a) 333, (b) 360.4 m s^{-1} ; 342 m s^{-1} . 2. 349 m s^{-1} . 3. 332 m s^{-1} . 4. 53.1° . 6. 680 Hz. 9. 1366 m. 11. 1115 Hz. 13. 1.83 s. 14. 6.7 m s^{-1} . 15. $6.44 \times 10^6 \text{ m s}^{-1}$. 16. 1174, 852 Hz. 17. (b) (i) 66 cm, (ii) (1) 550, (2) 545, (3) 600 Hz. 21. 10.8 db. 22. 25 W. 23. $5 \times 10^{-17} \text{ W m}^{-2}$.

EXERCISES 27 (p. 671)

1. (i) $\lambda/2$, (ii) $\lambda/4$, (iii) $\lambda/2$; 567 Hz. 2. 20 cm. 3. (a) 0.322, (b) 0.645 m. 4. (b) Amp.: (i) max, (ii) 0, (iii) max, (iv) half-max. 5. 267 Hz. 8. (i) 15.95 cm, (ii) 4.5 Hz. 9. $+5.2^\circ\text{C}$. 12. -3.2 , $+6.7\%$. 13. 239 Hz. 15. (a) 2 m, (b) 100 Hz. 16. touch $1/6\text{th}$ from end. 18. 2.08 kgf. 19. 100, 300, 500 Hz.

OPTICS

EXERCISES 28 (p. 685)

2. 39° . 3. 18.6° . 4. $47^\circ 10'$; $41^\circ 48'$ with vertical at oil surface. 8. 4.0×10^{-7} m.

EXERCISES 29 (p. 723)

2. 6.25×10^{-7} m. 3. 2.27×10^{-5} m. 4. 0.34 mm. 5. 1.5. 7. 1.64 mm; oil—centre bright, fringes closer. 8. 2.52×10^5 . 10. 1.42×10^{-6} m, receded. 11. $1\frac{1}{2}$. 13. inwards, 0.133 cm. 15. 2.32×10^{-7} m. 16. $2^\circ 35'$. 18. 3; 1.2 cm. 19. 6,800. 20. (b) 5 m s^{-1} . 21. 13.3° . 22. 3; 5.895×10^{-7} m. 23. 2.35 mm, 0.0785 rad. 25. 67.5° . 29. 49.0° . 32. 2.37 cm.

ELECTRICITY AND ATOMIC PHYSICS

EXERCISES 30 (p. 762)

2. 1.5×10^6 V. 4. (i) $E = \sqrt{3}q/4\pi\epsilon_0 z^2$, $V = q/2\pi\epsilon_0 z$, (ii) $E = q/4\pi\epsilon_0 z^2$ parallel to AB, $V = 0$. 7. (a) 9×10^4 N C $^{-1}$ (V m $^{-1}$), (b) 9000 V. 8. 1.59×10^{-19} C. 9. 3 electrons, 19.6 m s $^{-2}$ (2g). 12. 100 V. 13. 1.33×10^{-5} C, 6×10^5 V, yes.

EXERCISES 31 (p. 782)

1. (a) 20 V, (b) 2.5×10^{-3} J, originally 12.5×10^{-3} J. 2. 3 μ F; 20 V. 3. (a) 1.7×10^{-5} μ F, (b) 1.22×10^{-7} J, (c) A— 6.2×10^{-8} J; B— 6.0×10^{-8} J. 4. 33.9 m. 5. (a) 0.1 μ F, (b) 100. 6. 20 M Ω . 8. 3. 9. 5.5×10^{-4} , 0.0025×10^{-4} μ F. 10. 40,400 V m $^{-1}$. 11. 4 μ F. 12. 1.5×10^{-8} C, 2700 V. 13. 4.25×10^{-5} A. 14. (a) 0.988, (b) 0.88, (c) 440 V.

EXERCISES 32 (p. 805)

1. 18 Ω . 2. 3.6 mV. 3. 2.2 kJ kg $^{-1}$ K $^{-1}$ (2.2 J g $^{-1}$ K $^{-1}$). 4. 1.1 mm. 6. $\frac{1}{2}$, 1, 2 kW. 7. 36, 39%. 10. 200 cells, 0.30 g. 12. 18.75 kW.

EXERCISES 33 (p. 839)

1. 25, 50 V. 2. (a) 0.3 A, (b) 83×10^{-6} , 81×10^{-3} W. 3. $64n/(n^2 + 16)$, 8 A, $3/8 \Omega$. 4. $X = 2000 \Omega$; 1000 Ω . 5. -1% . 6. 2μ A div $^{-1}$. 7. 3.96×10^{-4} , 4×10^{-4} A. 8. $1\frac{2}{3}$ V. 9. V, 2.7 Ω . 10. 1 Ω . 11. (a) 45.9 cm from left hand, (b) 58.6 Ω parallel. 12. 5×10^{-4} A, 2.5×10^{-3} V; 1962 Ω . 13. 5.3×10^{-4} K $^{-1}$ (degC $^{-1}$). 14. 8.91 mV. 16. 2 Ω . 17. -5.1×10^{-4} K $^{-1}$ (degC $^{-1}$). 20. 3.23×10^{-4} s.

EXERCISES 34 (p. 871)

1. (a) 4.4×10^{-3} g, (b) 1.1×10^{-3} g, (c) 1860 s (31 min). 2. 5.1 degC. 3. 33×10^{-8} kg C $^{-1}$. 4. 1.49 V. 5. 1.7 V. 7. 23 Ω , 1.6 d, 50%. 8. 1.8 cm 3 . 9. 18.9 min. 10. 1.5 V.

EXERCISES 35 (p. 893)

1. 4.95 Ω , 20 mA. 2. $nABi/c$. 3. (a) shunt 0.0125 Ω , (b) series 9995 Ω . 4. 69 μ A. 5. AB, CD: 8.49×10^{-5} N horizontal; BC: 1.8×10^{-4} N at 20° to horizontal; 1.85×10^{-5} N m, (a) BC goes down, (b) BC goes down. 6. 2.1 A m $^{-2}$. 8. 0.0136 μ A/div. 9. series 140, 7490 Ω . 115 V.

EXERCISES 36 (p. 927)

3. 44.2 s $^{-1}$. 4. 4 div μ C $^{-1}$, 1.9 T (Wb m $^{-2}$). 6. 1.05 V. 8. 32.3 div. 10. 0.53 V. 11. 4 Ω , 8×10^{-3} H; 12.5 V. 12. (a) 4 A, (b) 2 A s $^{-1}$. 13. 0.93 s. 14. 75.4 sin ωt . 16. 1.6 mV. 18. 2.61×10^{-7} A, 1.02×10^{-11} W. 20. 1.05 V.

EXERCISES 37 (p. 946)

1. 8.7×10^{-3} V. 2. 6 cm from 12 A wire. 3. 3.64 A. 4. 8.4 mA. 6. $\mu_0 n I_1 I_2 l^2 / 2\pi d(d+l)$, 1.5×10^{-4} N. 8. $\mu_0 ix / \pi(x^2 + a^2)$; $x = a$.

EXERCISES 38 (p. 965)

1. 0.22 A. 2. 2500. 3. (i) 0.298 T (Wb m $^{-2}$), (ii) 238, (iii) 237.

EXERCISES 39 (p. 990)

1. 3.46 A. 2. (a) 0.1 A, (b) 3 J. 3. (a) 754 Ω , (b) 5.42 H. 4. 24 Ω . 5. 40 Ω , 0.24 H; 26.7 Hz, 7.1 V. 6. 1.126 μF ; 140°; 0.645. 7. X = 78 Ω , 7.8 V; Y = 99 Ω , 9.9 V. X: +46°, Y: -66°. 10. 71 V, 25/ π , 25 W, 25/ $\sqrt{3}$ W. 11. 1.57×10^{-4} V. 12. 0.06 H. 13. 4 Ω , 9.55 mH. 14. 0.4 V, 2 V m⁻¹, 3.2×10^{-19} N. 15. 1.3×10^{-12} T.

EXERCISES 40 (p. 1007)

2. 2.0006 V. 3. 14.4×10^{-3} m. 4. 0.5×10^{-4} T (Wb m⁻²). 5. 4e; changes to 2e and 3e. 6. (a) 1.5×10^{-6} m, (b) 1, (c) 5.0×10^{-4} m s⁻¹. 8. 2 cm. 9. 3.72×10^{-17} C. 10. 0.32 cm. 11. 1.8×10^{11} C kg⁻¹. 12. 1.6×10^{-19} C.

EXERCISES 41 (p. 1036)

7. 200. 10. 62000. 11. 1.8×10^{11} C kg⁻¹. 22. (i) 45, (ii) 13 k Ω .

EXERCISES 42 (p. 1065)

3. 9.6×10^5 min⁻¹. 4. β -particles emitted, 198. 8. 14.9 MeV. 9. 0.11. 10. (a) 4.23, (b) 4.16 MeV. 12. 56 s. 13. 3.36×10^{22} . 15. (a) 2.6×10^{12} :1, (b) 2.6×10^{13} m s⁻², (c) 9.6×10^{-17} J.

EXERCISES 43 (p. 1092)

1. 1.24×10^{-11} m. 2. 6.64×10^{-34} J s. 5. 2.3×10^{17} , 4×10^{-7} A. 6. 5000:1. 7. 1.8×10^{11} C kg⁻¹. 9. 3.1×10^4 Hz. 10. 8125 eV. 12. 2.0×10^{-7} m. 13. A, B, C, D = 6.5, 4.8, 4.3, 4.0×10^{-7} m. 14. (a) 9×10^{-7} m, (b) 7.8×10^5 m s⁻¹, (c) 1.7 V.

Index

- ABBE objective, 715
 refractometer, 434
Absolute expansivity, 280
 measurement, current, 940
 temperature, 224
 zero, 224
Absorption of radiation, 348–50, 359
 spectra, 464
Absorptive power (sound), 614
A.C. circuits, 970–9
Acceleration, 5, 87
 angular, 81
 in circle, 37
 of charge, 986
 of gravity, 7, 41, 66
Accommodation (eye), 508
Accumulator, 866, 870
Achromatic doublet, 515
 prisms, 459
Acoustics of rooms, 614
Adhesion, 133
Adiabatic change, 252
 curves, 252–3
 modulus, 162
A.f. amplifier (transistor), 1031–3
 (valve), 1016
Aerial, 986
Air breakdown, 749
Air-cell method, 432
Alcohol thermometer, 374
Alpha-particle, 1040, 1043–4
 , range of, 1043
 , scattering of, 1050
Alternating current, 912, 967
Alternators, 907
Ammeter, A.C., 967, 1014
 , D.C., 882
Ampere, 743, 793, 939
 balance, 940
 -hour, 869
 , circuit law of, 942
 -turns, 938
Amplification, transistor, 1031
 , valve, 1016
Amplitude, 44, 577, 607
 , A.C., 906, 968
ANDREWS' experiments, 314
Angle of contact, 132
Angular magnification, 526, 533
 momentum, 82
 velocity, 36
Anode (electrolysis), 844
 of valve, 1009
Anode characteristic, 1015
Anomalous expansion of water, 287
Antinode (e.m. wave), 984
 (sound), 642
Aperture, 402, 542
Apparent depth, 426
 expansivity, 280
APPLETON layer, 432
ARCHIMEDES' principle, 114
Armature, 907
ARRHENIUS' theory, 846
Artificial disintegration, 1060
Astigmatism, 511
Astronomical telescope, 533
Atmospheric pressure, 111
 refraction, 431
Atomic mass, 1053
 nucleus, 1050, 1052
 number, 1051, 1090
 spacing, 1072
 structure, 1051, 1090
 unit, 1057
Audibility, 609
Audio frequency, 606
AVOGADRO constant, 230, 993

B (flux density), 879, 920
 , measurement of, 920, 934
Back-e.m.f. of induction, 923
 of motor, 885
 of polarization, 857, 861
BAINBRIDGE spectrometer, 1057
Balance, common, 107
Balancing column method, 281
Ballistic galvanometer, 919
Balmer series, 1087
Band spectra, 462
Banking of track, 40
Bar, 112
Barometer, 113
 correction, 113, 286
Barrier p.d., 1026
Base (transistor), 1028

- Beats, 620
- BECQUEREL, 1040
- BERNOULLI's principle, 118
- BERTHELOT's method for γ , 212
- Beta-particles, 1040, 1043, 1045
- Bimetal strip, 274-6
- Binding energy, 1058
- Binoculars, 540
- BIOT and SAVART law, 931
- Biprism experiment, 691
- Black body, 351
 - radiation, 352-7
- BLACKETT, LORD, 1051
- BOHR's theory, 1084
- Boiling-point, 305
- Bolometer, 346
- BOLTZMANN constant, 235
- BOYLE's law, 220
- BOYS, G, 60
 - method (lens), 498
- BRAGG's law, 1059
- Breakdown potential, 749
- Breaking stress, 154
- BREWSTER's law, 720
- British Thermal Unit, 193
- Brittle, 154
- Brownian motion, 993
- Bubble chamber, 1049
- Bulk modulus, 161, 623
- BUNSEN's ice calorimeter, 209

- Calibration of thermometer, 190
- CALLENDAR and MOSS, 282
- Calorie, 193
- Calorimeter, 200
- Camera lens, 542
- Candela, 562
- Candle-power, 562
- Capacitance, 767-775, 776
 - , measurement of large, 768
 - , measurement of small, 775
- Capacitor, charging of, 765, 970
 - , discharging of, 766
- Capacitor, electrolytic, 854
 - mica, 765
 - paper, 765
 - parallel-plate, 771
 - variable, 770
- Capacitors in parallel, 776
 - in series, 777
- Capillarity, 131, 139
- CASSEGRAIN, 545
- Cathode (valve), 1009
- Cathode-ray oscillograph, 1011
- Cathode rays, 996, 999
- Cells, primary, 785, 858
 - , secondary, 866
 - , series and parallel, 801
- CELSIUS temperature scale, 190

- Centre of gravity, 103
 - of mass, 24, 103
- Centrifuge, 38
- Centripetal force, 38
- CHADWICK, 1052
- Chain reaction, 1062
- Characteristics of sound, 606
- Charge carriers, 891
- Charge on conductor, 739, 742
 - on electron, 732, 994
- CHARLES' law, 222
- Chemical activity, 1090
 - equivalent, 845
- CHLADNI's figures, 669
- Chromatic aberration, 513
- Circle, motion in, 36-41
- Circular coil, 876, 932, 935
- CLÉMENT and DÉSORMES, 255
- Clinical thermometer, 373
- Clock, loss and gain of, 276
- Closed pipe, 645
- Cloud chamber, 1048
- COCKCROFT-WALTON, 1060
- Coefficient of friction, 171-2
 - of restitution, 22
 - of viscosity, 175
- Coercive force, 954
- Coherent sources, 687
- Cohesion, 133
- Collector (transistor), 1028
- Collimator, 444
- Colours of sunlight, 454
 - of thin films, 699
- Combination of thin lenses, 499
- Common balance, 107
 - base circuit, 1029
 - emitter circuit, 1030
- Comparator method, 270
- Comparison of e.m.f.s., 818
 - of resistance, 824
- Compensated balance-wheel, 276
- pendulum, 277
- Components of force, 97
 - of velocity, 8
 - resolved, 8
- Compound microscope, 529
 - pendulum, 90
- Concave lens, 478
 - mirror, 402
- Concentric spherical capacitor, 770
- Condensing lens, 541
- Conduction in metals, 786
 - through gases (electrical), 998
- Conduction, thermal, 330
 - comparison of, 340
 - of bad conductor, 334, 336
 - of good conductor, 335
 - of liquids, 338
- Conductivity, 856
- Conductors, 786, 788

- Conjugate foci (points), 490
- Conservation of energy, 28, 195
 - of momentum, 18, 82
- Conservative forces, 28
- Constant pressure experiment, 222
 - volume gas scale, 366
 - thermometer, 367
- Continuous flow calorimeter, 206
 - spectra, 463
- Convection, 359
- Convex (converging) lens, 478
 - mirror, 402
- Cooling, correction for, 203
 - NEWTON's law of, 201
 - specific heat capacity by, 205
- Copper oxide rectifier, 1011
- Corkscrew rule, 876
- Corona discharge, 749
- Corpuscular theory, 679
- Cosmic rays, 999
- Coulomb, 743
- Couple on coil, 101, 885
 - on rigid body, 80
- Couples and work, 101-2
- Couplings, circuit, 1018
- COUDÉ, 545
- Critical angle, 430
 - potential, 1083
 - temperature (gas), 243, 316
 - velocity (viscosity), 177
- CROOKES' dark space, 997
- Crystal structure, 1071
- Cubic expansivity, gas, 223
 - liquid, 279
 - powder, 284
 - solid, 278
- CURIE-JOLIOT, 1052
- CURIE temperature, 963
- Current balance, 879, 940
 - , electron, 786
- Current and units, 846, 939
- Current, potentiometer, 822
- Curvature, 683
- Curved mirror formulae, 407-411
- Cylindrical surface, 981

- DALTON's law of partial pressures, 221
- Damped oscillation, 46, 580
- Damping of galvanometer, 580, 913
- DANIELL cell, 860
- DE BROGLIE's law, 1075
- Decay current, 1046
- Declination, 945
- Decibel, 608
- Defects of vision, 508
- Degrees of freedom, 238-9
- Dekatron counter, 1041
- Demagnetization, 958
- Density, 114
 - , determination of, 114-6
 - of earth, 62
 - of liquid, and temperature, 280
 - of water, 288
- Depth of field, 543
- Detection, diode, 1022-3
- Deuterium, 1056
- Deviation by mirror, 392-5
 - by prism, 442, 681
 - by sphere, 474
 - by thin lens, 481
 - by thin prism, 456
- Deviations from gas laws, 127, 242, 313-8
- DEWAR flask, 351
- Dew-point, 310
 - hygrometer, 315
- Diamagnetism, 962
- Dielectric, 772
 - constant, 772
 - polarization, 773
 - strength, 772
- Difference of specific heats, 244
- Differential expansion, 273
 - steam calorimeter, 246
- Diffraction at lens, 704
 - at slit, 58, 702
 - , electron, 1074
 - , microwave, 988
 - , X-ray, 1069
- Diffraction grating, 707
- Diffuse reflection, 391
- Diffusion, 236
- Dilatometer, 285
- Dimensions, 30
 - applications of, 31, 162, 176, 181
- Diode valve, 1009, 1022
- Dioptre, 684
- Dip, 944
- Dipole (electric), 773
- Direct-vision spectroscope, 461
- Discharge, capacitor, 766
 - tube, 996
- Disintegration, nuclear, 1052, 1060
- Dislocation, 155
- Dispersion by prism, 457, 681
- Dispersive power, 458
- Displacement (waves), 577, 583, 585
 - of lens, 491, 495
 - by rectangular block, 427
- Dissociation, electrolytic, 847
- Distance-time curve, 2
- Diverging lens, 478
- Domains, magnetic, 951, 963
- DOPPLER effect, 627
- Double refraction, 720
- Drift velocity, 787
- Driving mirror, 406
- Drops, formation of, 1048
- Dry cell, 865
- Ductile, 154

- DULONG and PETIT, 280
 Dust-tube experiment, 666-9
 Dynamo, 905
 Dyne, 14
- Ear, structure of, 594
 Earth, density of, 62
 , escape velocity, 68
 mass of, 62
 potential of, 752
 Earth's horizontal component, 922, 945
 magnetic field, 944
 vertical component, 921, 945
 Echo, 614
 Eclipses, 387
 Eddy currents, 913
 Efficiency (electrical), 798, 869
 EINSTEIN's mass law, 29, 1057
 photon theory, 1078
 Elastic collisions, 20
 limit, 153
 Elasticity, 153
 , adiabatic, 162, 624
 , isothermal, 162, 623
 , modulus of, 155, 623
 Electric field, 744, 750, 754
 , moving, 981
 intensity, 744, 754
 potential, 750-60
 Electrical calorimetry, 206, 213, 248
 Electrochemical equivalent, 845
 Electrode potential, 862
 Electrodes, 844
 Electrolysis, 844
 back e.m.f., 856
 , FARADAY's laws of, 845, 849
 Electrolyte, 844, 855
 Electromagnetic induction, 895
 radiation, 986
 waves, 980
 Electromagnetism, 875, 931
 Electron charge, 732, 995, 1000
 diffraction, 1073
 , e/m , 1002
 lens, 1013
 mass, 1004
 motion, 787, 1001-5
 orbit, 848, 1084, 1089
 shells, 1089
 -volt, 756
 Electrons, 732, 786, 996
 Electrophorus, 736
 Electroscope, 733, 759
 Electrostatic fields, 744, 750
 intensity, 744, 754
 shielding, 760
 Emitter (transistor), 1028
 E.M.F., 794, 901, 906
 , determination, 818, 825
 Emission spectra, 462, 1088
- Emissivity, 356
 Emulsion detector, 1050
 End-correction, 650, 653
 Energy and matter, 29, 1057
 Energy, conservation of, 29, 195
 , electrical, 750, 790
 exchanges, 52, 578
 in capacitor, 779, 926
 in coil, 925
 levels (atom), 1081-8
 luminous, 561
 mechanical, 25, 28
 nuclear, 1061, 1062
 shells, 1089
 sound, 607
 Energy in wire, 888
 , kinetic, 25
 , potential, 26
 rotational, 75, 86, 238
 translational, 235
 vibrational, 52
 Equation of state, 226
 Equilibrium, conditions of, 97, 99, 100
 Equipotential, 756-8
 Equivalent valve circuit, 1017
 Erg, 25
 Errors, counting, 1042
 Escape velocity, 68
 Evaporation, 294, 297
 Excess pressure in bubble, 137, 144
 Excitation potential, 1087
 Expansion, cubic, 278
 , linear, 269
 of hole, 273
 of scale, 273
 of water, 287
 superficial, 278
 External work, 229, 244
 Extraordinary ray, 719
 Eye, structure of, 389, 508
 Eyepiece, HUYGHENS', 519
 , RAMSDEN's, 519
 Eye-ring, 532, 534, 538
- Falling-plate experiment, 600
 Falling sphere, viscosity by, 182
 Far point of eye, 508
 sight, 570
 Farad, 768
 FARADAY's dark space, 997
 ice-pail experiment, 738
 laws of electrolysis, 845, 849
 law of induction, 898
 Feedback, energy, 1030
 Ferromagnetism, 874, 952, 956, 963
 Field of view of microscope, 532
 of mirror, 406
 of telescope, 534, 538
 Filter pump, 257
 Fine beam tube, 1004

- Fission, nuclear, 1061
- Fixed points, 190, 366
- FIZEAU's method, 555
- Flame, sensitive, 650
- FLEMING's left-hand rule, 878
 - right-hand rule, 900
- Flotation, 117
- Flow, laminar, 177
 - turbulent, 177
- Flux, electric, 745-8
 - magnetic, 898
- Flux density, B , 879, 885, 920, 934, 948
 - linkages, 898, 918
 - f -number, 542
- Focal length of lenses, 479, 482, 494-501
 - of mirrors, 404, 413
- Focusing, electron, 1013
- Foot-candle, 562
- Force, 13
 - between currents, 939
 - due to expansion, 272
 - on charges, 881, 902
 - on conductor, 878
- FOUCAULT's method, 556
- FRANCK-HERTZ, 1081
- FRAUNHOFER's lines, 464
- Free oscillation, 46
- Freezing, 295, 297, 321
- Frequency, 577
 - by falling plate, 600
 - by resonance tube, 654
 - by sonometer, 660
 - by stroboscope, 602
 - fundamental, 646, 648, 658
- FRESNEL's biprism, 691
- Friction, coefficient of, 171-2
 - laws of, 173
 - limiting, 171
- Fusion, nuclear, 1062
- Galilean telescope, 537
- GALILEO, 49, 537
- Galvanometer, 882
 - resistance, 836
 - sensitivity of, 836
- Gamma-rays, 1043
 - inverse-square law, 1045
 - range, 1043
- Gas, 318
 - constant, 228-30
 - equation, 227
 - ideal, 127, 231
 - laws, 220-8
 - molar, 230
 - real, 127, 313
 - thermometer, 367
 - velocity of sound in, 668
- Gases, diffusion of, 236
- Gas-lens, 615
- Gaseous state, 127, 318
- GAUSS's theorem, 747
- GEIGER and MARSDEN, 1050
 - MÜLLER (G M) tube, 1040
- Generators, 907, 910
- Germanium, 1024
- Glancing angle, 392
- Gold-leaf electroscope, 733, 759
- GRAHAM's law, 236
- Gravitation, law of, 59
- Gravitational constant, 60
 - mass, 62
 - orbits, 63, 69
 - potential, 68
- Gravity, acceleration of, 7, 41, 49, 66
 - motion under, 7
- Grazing incidence, 448-9
- Grid (valve), 1015
 - bias, 1017
- Ground state, 1083
- Gyration, radius of, 78
- Half-life, 1046
- Half-wave rectification, 1010
- Hall voltage, 891
- Harmonics of pipes, 649
 - of strings, 658
- HARRISON's pendulum, 277
- Heat capacity, 199
 - energy, 193
- Heating effect of current, 206, 790
- Heavy hydrogen, 1056
- HEBB's method, 618
- Helium, 1044, 1063
- HELMHOLTZ coils, 936
 - resonator, 611
- Henry, the, 924
- High temperature measurement, 375
 - tension transmission, 793
- Holes, 1024
- Hollow conductor, 759
- Homopolar generator, 911
- HOOKE's law, 154
- HOPE's experiment, 289
- Horizontal component, B_H , 922, 945
- Horse-power, 25
- Hot-wire ammeter, 968
- Humidity, 310
- HUYGHENS construction, 676
 - eyepiece, 519
- Hydrogen atom, 1051, 1085
 - isotopes, 1056
- Hydrogen gas thermometer, 367
- Hydrometers, 117
- Hygrometry, 309-12
- Hysteresis, 953-8
- Ice calorimeter, 209
 - pail experiment, 738
 - point, 190

- Illumination, 562
 - meter, 571
- Images in curved mirrors, 405-12
 - in lenses, 485-7
 - in plane mirror, 396
- Impedance, 975
- Impulse, 16
- Impurity, metal, 789
 - , semiconductor, 1025
- Inclined mirrors, 394, 399
- Induced charge, 734-9
 - current, 895
 - e.m.f., 896, 899
- Inductance, 924
- Induction coil, 903
 - , electrostatic, 734
 - , magnetic, 941
- Inelastic collisions, 20
- Inertia, 12
 - , moments of, 75-9
- Inertial mass, 62
- Infra-red radiation, 344, 456
 - spectrometer, 348
- INGENHOUSZ's experiment, 340
- Insulators, 731
- Intensity, electric, 744, 754
- Intensity of magnetization, 949
 - of sound, 607
- Interference (light), 687
 - in thin film, 697
 - in wedge films, 692
 - of waves, 592, 616
- Intermolecular forces, 125
 - energy, 125
- Internal energy of gas, 237-241
 - work, 241, 306
- Internal resistance of cell, 794, 797, 820
 - of valve, 1016, 1018
- International temperature scale, 368
- Invar, 270, 274
- Inverse-square law (electrostatics), 743
 - (light), 565
- Inversion temperature, 324
- Ionization current, 998
 - of gases, 998
 - potential, 1088
- Ions, 848, 998
- Isothermal bulk modulus, 162
 - change, 249
 - curves, 250, 314-7
 - work, 251
- Isotopes, 1056
- JAEGER's method, 144
- JOLY's steam calorimeter, 247
- JOULE-KELVIN effect, 322
- Joule, the, 18, 198
- JOULE heating, 790
- Junction diode, 1027
- KELVIN, 791, 920
- KELVIN scale, 190, 224
- KEPLER's laws, 58, 85
- Kilocalorie, 196
- Kilogramme (kg), 13
 - force (kgf), 14
- Kilowatt, 792
- Kilowatt-hour, 794
- Kinetic energy, 25
 - theory of gases, 231-40
 - of liquids, 304
- KIRCHHOFF's law (electricity), 827
 - (light), 464
- KUNDT's tube, 666
 - determination of γ , 668
 - velocity in gas, 668
 - velocity in rod, 667
- LAMBERT's law, 566
- Laminar flow, 174, 177
- Laminations, 913
- LAPLACE's correction, 624
- Latent heat, 208-13
 - of evaporation, 211, 294, 306
 - of fusion, 208, 294
- Lateral inversion, 397
 - magnification, 487
- Laue diffraction, 1069
- Leakage current, 1002
- LECLANCHÉ cell, 864
- LEES' disc method, 336
- Lens, 471, 478
 - condensing, 541
 - conjugate foci of, 490
 - formula, 482
 - of eye, 389
- LESLIE cube, 350
- LENZ's law, 896
- Limiting angle of prism, 449
- LINDE's process, 323
- Line spectra, 463
- Linear flow (heat), 330
 - expansivity, 269
- Lines of force, electric, 741, 745, 757
 - , magnetic, 875-7
- Liquefaction of gases, 321-4
- Liquid state, 294
- LLOYD's mirror, 697
- Long sight, 509
- Longitudinal waves, 56, 584
- Loudness, 607
- Loudspeaker, 596
- Lumen, 562
- Luminance, 567
- Luminous flux, 562
 - intensity, 563
- LUMMER-BRODHUN photometer, 568
- Lux, 562
- Lyman series, 1087

- MACLEOD gauge, 259
- Magnetic circuit, 959
 - cycle, 955
 - fields, 874-7
 - ; moving, 980
 - flux, 898
 - flux-density, 880, 898, 948
 - induction, 948
 - intensity, 941
 - coil, 932, 935
 - solenoid, 937
 - straight wire, 936
 - materials, 948
 - meridian, 945
 - moment, 885
 - pole, 874
 - susceptibility, 951
- Magnetometer, vibration, 886
- Magnetomotive force, 960
- Magnification, angular, 526
 - lateral (linear), 409, 487
 - of microscope, 527, 530
 - of telescope, 533-9
- Magnifying glass, 527
- Mains frequency, 967
- Majority carriers, 1025
- Mass-energy relation, 29, 1057
 - number, 1053
 - spectrometers, 1056-7
- Maximum power, 800
 - thermometer, 374
- MAXWELL, 231, 876
- Mean square velocity, 233
 - velocity, 233
- Mechanical equivalent of heat, 194, 197
- Megelectron-volt, 1058
- MELDE's experiment, 661
- Melting-point, 295
 - effect of pressure, 296
- Mercury, absolute expansion, 281
 - thermometer, 189, 372
- Metal-glass seal, 277
- Metal conductor, 56
 - rectifier, 1011
- Metre bridge, 835
- Mica capacitor, 765
- MICHELSON's method (light), 559
- Microfarad, 768
- Microphone, carbon, 595
 - ribbon, 596
- Microscope, compound, 529
 - simple, 527
- Microwaves, 988-9
- MILLIKAN's experiment, *e*, 994
 - photoelectric, 1080
- Minimum deviation by prism, 442, 681
 - distance, object-image, 489
- Mirage, 431
- Mirror galvanometer, 882
- Modulation, 1022
- Modulus of elasticity, adiabatic, 162, 624
 - bulk, 161
 - isothermal, 162
 - of gases, 162
 - of rigidity, 163
 - YOUNG's modulus, 156
- Molar gas constant, 230
- Mole, 230
- Molecular forces, 125-7, 242, 322
 - sphere of attraction, 128
- Moment, magnetic, 885, 887
- Moment of couple, 101, 885
 - of force, 98
- Moments of inertia, 75-90
 - measurement, 88, 90
 - theorems of, 78-9
- Momentum, angular, 82
 - linear, 16
 - conservation of angular, 82
 - of linear, 18
- Moon, motion of, 59
- MOSELEY's law, 1073
- Motion, NEWTON's laws of, 12
 - in circle, 36-41
 - in straight line, 1-7
 - of projectile, 9
 - simple harmonic, 43-54
 - under gravity, 7
- Motors, 884
- Mount Palomar telescope, 544
- Moving-coil ammeter, 882
 - galvanometer, 882
 - loudspeaker, 596
 - voltmeter, 811, 814
- Moving-iron meter, 967
- Multimeters, 813
- Multiple images, 423
- Multiplier, 812
- Musical interval, 606
- Mutual characteristic, 1015
 - conductance, 1016
 - induction, 927
- Near point of eye, 508
- Negative charge, 731
- NERNST's calorimeter, 208
 - theory, 862
- NEUMANN's law, 898
- Neutral equilibrium, 106
 - temperature, 804
- Neutron, 1052, 1061
- NEWTON's law of cooling, 201
 - laws of motion, 12
 - lens formula, 490
 - rings, 693
 - velocity (sound) formula, 624
- Newton, the, 13
- NICOL prism, 722
- NIFE cell, 870
- No parallax, 398

- Nodes, 590
 - in air, 642, 646, 649
 - in pipes, 646
 - in rod, 666
 - in strings, 657
- Non-conservative forces, 28
- Non-inductive resistance, 924
- Non-ohmic conductors, 790
- N-p-n transistor, 1028
- N-semiconductor, 1025
- Nuclear charge, 848, 1051, 1073
 - dimension, 1051
 - reactions, 1059-64
 - stability, 1059
 - structure, 1052
- Nucleon, 1058
- Nucleus, 1050
- Objective, lantern, 541
 - , microscope, 530
 - , telescope, 533
- Octave, 606
- OERSTED, 875
- Ohm-meter, 814
- OHM's law (electricity), 787, 789, 797
 - electrolytes, 854
- Oil-immersion objective, 715
- Open pipe, 647
- Opera glasses, 538
- Optical instruments, 525
 - lever, 393
 - path, 698
 - pyrometer, 376
 - spectra, 462, 1084, 1087
- Ordinary ray, 719
- Oscillation, damped, 46, 580
 - , electric, 578
 - , energy exchanges, 52
 - , free, 46
 - of liquid, 52
 - of magnet, 886
 - of rigid body, 89
 - of spring, 50
 - , undamped, 46
- Oscillator, transistor, 1034
 - triode, 1035
- Oscillatory current, 1019
- OSTWALD viscometer, 180
- Overtones, 611
 - of pipes, 646, 648
 - of strings, 658
- Oxygen, liquefaction of, 324
- Paper capacitor, 765
- Parallel-plate capacitor, 755, 771
- Parabolic mirror, 403
- Parallel axes, theorem of, 78
 - forces, 99
- Parallelogram of forces, 96
 - of velocities, 8
- Paramagnetism, 963
- Parking orbit, 64
- Particle, nature, 1076
 - wave duality, 1082
- PAULI's Exclusion Principle, 1090
- Peak value, 968
- PELTIER effect, 791
- Pendulum, compensated, 277
 - compound, 91
 - simple, 48
- Period, 44
- Permeability, 951
 - , relative, 951
 - , vacuum, 940
- Permittivity, 743-4
 - relative, 772
 - , vacuum, 743
- Perrin tube, 1000
- Phase angle, 579, 971
 - difference, 579, 971
- Phon, 609
- Photon, 1079
- Photo-cells, 1078-9
- Photoelectricity, 1077
- Photometry, 561
- Phototransistor, 1035
- Pinhole camera, 542
- Pipe, flow through, 118
- Pitch, 606
- PLANCK constant, 1076, 1080
- Plane conductor, 748
 - mirror, 391
- Plane-progressive wave, 56, 584
- Planetary motion, 58
- Plasma, 1064
- PLATEAU's spherule, 129
- Plates, vibrations of, 669
- Platinum resistance thermometer, 191, 369
- P-n junction, 1026
- P-n-p transistor, 1028
- Points, action at, 737, 749
- Poise, 175
- POISEUILLE's formula, 176
- POISSON's ratio, 166
- Polarization (e-m wave), 987
 - (light), 716
 - by double refraction, 720
 - by reflection, 720
 - by tourmaline, 717
- Polarization of cell, 862
 - of dielectric, 773
- Polarizing angle, 720
- Polar molecules, 773
- Pole of magnet, 874
 - of mirror, 402
- Positive charge, 731
 - column, 997
 - ion, 847, 1054
 - rays, 1054

- Post Office box, 834
- Potential difference, 751, 791, 793
 - divider, 810
- Potential, electric, 780-90
- Potential energy (mechanical), 26
- Potentiometer, 817-25
- Powder expansion, 284
- Power (electrical), 792, 799, 880
 - (A.C.), 978
 - , mechanical, 25
 - of lens, 683
 - of surface, 473
- Presbyopia, 510
- Pressure, atmospheric, 111
 - coefficient, 225
 - curves (sound), 585, 644
 - , gas, 231
 - liquid, 109
 - standard, 112
- PRÉVOST's theory, 355
- Principal focus, 479
 - section, 720
- Principle of Superposition, 588, 616
- Prism, 441
 - binoculars, 540
 - deviation by, 442
 - limiting angle of, 449
 - minimum deviation by, 443
 - NICOL, 722
 - thin, 456
 - total reflecting, 449
- Progressive wave, 57, 584
- Projectiles, 9
- Projection lantern, 541
- Proton, 1051
- P-semiconductor, 1026
- PULFRICH refractometer, 433
- Pump, 257
- Pure spectrum, 455
- Pyrometry, 375-9
- Quality of sound, 611
- Quantity of charge, 743, 767, 845
 - of heat, 194
- Quantization of energy, 1082
- Quantum theory, 1074, 1080, 1090
 - of light, 1080
- Quenching agent, 1041
- QUINCKE's tube, 617
- Radial field, 882
 - flow, 339
- Radiation (thermal), 343-60
 - laws, 346, 354
 - pyrometer, 375
 - wavelengths, 345
- Radioactivity, 1040
- Radio-frequency, 1021
- Radium, 1053
- Radius of curvature, lens, 498
 - of gyration, 78
 - of mirrors, 404, 413-6
- Rainbow, 475
- RAMSDEN's eyepiece, 519
- Range (radioactivity), 1043
- Rarefaction, 585
- Ratemeter, 1042
- Ratio of specific heats, 245, 255
- RAYLEIGH potentiometer, 825
- Reactance of capacitor, 972
 - coil, 973
- Real gases, 253, 313
 - image, 397, 405, 486
- Rectification, 1010
- Reflection of light, 391
 - of sound, 613
 - factor, 568
- Refraction at plane surface, 420, 679
 - at spherical surface, 471-5
 - through prism, 441, 681
 - wave theory of, 679
- Refraction of sound, 615
- Refractive index, 420, 679
 - absolute, 421
 - of glass, 428
 - of glass prism, 444
 - of lens, 482, 513
 - of liquid, 425, 432, 500
- Refrigeration, 318
- REGNAULT, 281
 - hygrometer, 311
- Relative density, 114
 - humidity, 310
 - permeability, 951
 - velocity, 10
- Reluctance, magnetic, 960
- Remanence, 953
- Resistance, 790, 824, 829
 - low, 824
 - of valve, 1016
- Resistance box, 809
- Resistance thermometer, 192, 369
- Resistances in parallel, 808
 - in series, 807
- Resistivity, 788
- Resolution of forces, 97
- Resolving power, microscope, 715
 - telescope, 705
- Resonance (sound), 582, 651
 - tube, 653
- Resonance, series, 977
- Resonant frequency, 582, 977
- Resonators, 611
- Rest mass, 30
- Retardation, 6
- Retina, 389
- Reverberation, 614
- Reversibility of light, 390
- Reversible change, 251-2

- REYNOLDS, 177
 Rigid body, motion of, 75-90
 Rigidity, modulus of, 163
 Rocket motion, 22
 Rod, longitudinal vibration in, 665
 , moment of inertia, 76
 ROENTGEN, 1019
 Root-mean-square current, 968
 velocity, 234
 Rotary pump, 258
 Rotation of rigid body, 73
 Rotational energy, 75
 RUMFORD, 194
 RUTHERFORD, 1044, 1050
 and ROYDS, 1044
 RYDBERG constant, 1087
- SABINE, 614
 Satellites, 59, 63
 Saturated vapours, 297-307
 Saturation current, diode, 1010
 , gas, 998
 Scalars, 2
 Scaler, 1041
 Scales of temperature, 190, 366-71
 Scintillation photomultiplier, 1049
 Screening, 760
 Search coil, 920
 SEARLE's method, 335
 Secondary cells, 866
 coil, 908
 ionization, 1049
 SEEBECK effect, 803
 Selenium cell, 1077
 Self-induction, 922
 Semiconductors, 891, 1023
 Sensitive flame, 650
 Sensitivity of balance, 107
 meter, 837
 Series A.C. circuits, 974
 Series-wound motors, 917
 Sextant, 395
 Shear modulus, 163
 -strain, 163
 stress, 163
 Shells, energy, 1090
 Shunt, 812
 -wound motors, 917
 SI units, 1096-1100
 Sign convention, 407, 473
 Silicon, 1024
 Simple cell, 859
 harmonic motion, 43-7
 pendulum, 48
 Slide-wire bridge, 835
 Soap film, 136
 Solenoid, 877, 938
 Solid angle, 562
 friction, 171
 state, 294
 detector, 1041
 Solid, velocity in, 56, 162, 666
 Solution pressure, 862
 Sonometer, 659
 Sound track film, 599
 Sound waves, 585, 606
 Specific heat capacity, 199
 , constant pressure, 244, 248
 , constant volume, 243, 246
 of liquid, 200, 205, 206
 of solid, 200, 208
 of water, 199, 206
 Specific heats, difference of, 245
 Spectra, 462, 1087, 1089
 , hydrogen, 1087
 Spectrometer, 444
 , X-ray, 1073
 Spectroscope, direct-vision, 461
 Spectroscopy, 463
 Spectrum, hydrogen, 462
 pure, 455
 sunlight, 454, 464
 Speed in circle, 36
 Sphere capacitance, 770
 , field due to, 747
 Spherical aberration, 518
 mirrors, 402
 Spin, electron, 948, 1090
 Spring, spiral, 50-1
 Stability, 105
 nuclear, 1059
 Standard cell, 865
 Starting resistance, motor, 915
 States of matter, 125
 Stationary (standing) waves, 57, 589
 pipe, 646, 648
 rod, 666
 string, 657
 transmission line, 984
 Steam calorimeter, 247
 point, 190
 STEFAN constant, 355
 STOKES' law, 181
 Straight conductor, 876, 933-5
 Strain, bulk, 161
 , shear, 163
 , tensile, 155
 , torsional, 163
 Streamlines, 118, 174
 Stress, bulk, 161
 , shear, 163
 tensile, 155
 , torsional, 163
 Stretched wire energy, 158
 Strings, harmonics in, 658
 resonance in, 657
 Stroboscope, 589, 602
 Substitution method, 836
 Superficial expansion, 278

- Supercooling, 295
- Surface density, 742, 760
- Surface energy, 146
- Surface tension, 128
 - and excess pressure, 137
 - dimensions of, 130
 - measurement of, 134-6, 144
 - theory of, 128
 - variation of, 145
- Susceptibility, 951
- Telephone earpiece, 596, 956
- Telescope, astronomical, 533
 - , galilean, 537
 - , reflecting, 544
 - , terrestrial, 536
- Temperature, 189
 - coefficient, 837
 - gradient, 331, 334
- Tensile strain, 155
 - stress, 155
- Terminal velocity, 181
- Terrestrial magnetism, 945
- Tesla (T), the, 880, 898
- Thermal capacity, 199
 - conductivity, 331
 - insulators, 335
- Thermistor, 1034
- Thermocouple, 191, 371, 803, 825
- Thermodynamic scale, 190
- Thermoelectricity, 803, 826
- Thermoelectric thermometer, 370
- Thermometers, 191, 367-75
- Thermonuclear reaction, 1063
- Thermopile, 346
- Thermos flask, 351
- Thermostat, 274
- Thin prism, deviation by, 456
 - dispersion by, 457
- THOMSON effect, 804
 - experiment, 1002, 1054
- Timbre, 611
- Threshold of hearing, 609
- Time-base (C.R.O.), 1012
- Toroid, 938
- Torque, 80, 101
- Torricelli vacuum, 112
 - theorem, 120
- Torsion, 164
- Torsional vibration, 165
- Total internal reflection, 430
 - radiation pyrometer, 377
 - reflecting prism, 449, 540
- Tourmaline, 717
- Transformers, 908
- Transistor, 1028
 - amplifier, 1029-33
 - coupling, 1018
- Translational energy, 235
- Transmission line, 983
- Transmutation, 1061
- Transverse waves, 584, 663
- Triangle of forces, 97
- Triode, 1015
 - amplifier, 1016
 - characteristics, 1015
 - constants, 1016
 - detector, 1023
 - oscillator, 1020
- Triple point, 190, 319
- Tritium, 1064
- True expansivity, 280
- Tuning, 978
- Tuning fork, 600, 610
 - , frequency of, 600-3
- Turbulent motion, 177
- Ultra-violet rays, 456
- Ultrasonics, 625
- Uniform acceleration, 5
 - velocity, 1
- Universal gas constant, 230
- Unstable equilibrium, 105
- Unsaturated vapour, 297, 307
- Upthrust, fluid, 114
- Uranium, 1040
 - fission, 1061
 - series, 1053
- Valency, 845
- Valve, thermionic, 1009, 1015
- VAN DE GRAAFF generator, 737
- VAN DER WAALS', 127, 242
- Vapour pressure, 297
- Variable capacitor, 768
- Variation of g , 41, 66
- Variation, magnetic, 945
- Vectors, 1, 9, 774
- Velocities, addition of, 9
 - , subtraction of, 10
- Velocity, angular, 36
 - critical, 177
 - , relative, 10
 - , terminal, 181
 - time curve, 3
 - , uniform, 1
- Velocity of light, 388, 551
 - BRADLEY, 554
 - FIZEAU, 555
 - FOUCAULT, 556
 - MICHELSON, 559
 - ROEMER, 551
- Velocity of sound in air, 587, 618, 653
 - in gas, 163, 623
 - in pipe, 653
 - in rod, 163, 666
 - in string, 657
 - in water, 163, 620, 623
- Vibrating reed switch, 775
- Vibration magnetometer, 886

- Vibrational energy, 52
- Vibrations, forced, 581, 651
 - in pipes, 645
 - in plates, 669
 - in strings, 656
 - longitudinal, 584
 - resonant, 582; 651
 - transverse, 584, 656, 717
- Virtual image, 397, 406, 486
 - object, 412, 497
- Viscosity, 174
 - determination of, 179, 182
 - dimensions of, 175
 - falling sphere and, 181
 - , gas, 182
 - variation of, 181
- Vision, binocular, 389
 - defective, 508
- Visual angle, 525
- Volt, 750, 792
- Voltage amplification, 1033
- Voltmeter, 844
- Voltmeter, moving-coil, 811, 814, 821
- Volume expansivity, gas, 223
 - , liquid, 279
- Watt, 25
- Wattmeter, 886
- Wave, 55, 587
 - equation, 55, 587, 591
 - nature (particle), 1073
 - theory, 675
- Waveform analysis, 610
- Wavefront, 676
- Wavelength of light, 690, 710
 - of sound, 616, 646
- Weber, the, 898
- Weighing correction, 107-8
- Weight thermometer, 283
- Weightlessness, 65
- WESTON cadmium cell, 865
- Wetting surfaces, 133
- WHEATSTONE bridge, 829
- Whispering gallery, 614
- WIEN's law, 355
- WILSON cloud chamber, 1048
- Wind, effect of, 615
- Work, 24
 - done by couple, 102
 - done by gas, 229
 - in stretching wire, 158
- Work function, 1048
- X-rays, 1068
 - , diffraction of, 1070
 - spectra, 1069, 1073
 - spectrometer, 1071
- Yield point, 154
- YOUNG's experiment, 688
 - modulus, 156
 - , determination of, 156
 - , dimensions of, 156
- Zener diode, 1027
- Zero potential (atom), 1085, 1089
 - (electrical), 752
 - (gravitational), 68

M. Nelkon & P. Parker

Advanced Level Physics

